Analysis of Adaptive Multilevel Splitting algorithms in an idealized case

Charles-Edouard Bréhier, Tony Lelièvre, Mathias Rousset

ANALYSIS OF ADAPTIVE MULTILEVEL SPLITTING ALGORITHMS IN AN IDEALIZED CASE

CHARLES-EDOUARD BRÉHIER, TONY LELIÈVRE, AND MATHIAS ROUSSET

Abstract. The Adaptive Multilevel Splitting algorithm [4] is a very powerful and versatile method to estimate rare events probabilities. It is an iterative procedure on an interacting particle system, where at each step, the k less well-adapted particles among n are killed while k new better adapted particles are resampled according to a conditional law. We analyze the algorithm in the idealized setting of an exact resampling and prove that the estimator of the rare event probability is unbiased whatever k. We also obtain a precise asymptotic expansion for the variance of the estimator and the cost of the algorithm in the large n limit, for a fixed k.

Contents

1. Introduction

Let X be a real random variable, such that $X > 0$ almost surely. We want to approximate the following probability:

$$p = \mathbb{P}(X \geq a), \tag{1}$$

1991 *Mathematics Subject Classification.* 65C05; 65C35; 62G30.
Key words and phrases. Monte-Carlo simulation, rare events, multilevel splitting.

where $a > 0$ is a given threshold, such that $p > 0$. When a goes to infinity, the above probability goes to 0, meaning that $\{X \geq a\}$ becomes a rare event. Such problems appear in many contexts, such as molecular dynamics simulations [6] or reliability problems with many industrial applications, for example.

Estimating a rare event using a direct Monte Carlo estimation is inefficient, as can be seen by the analysis of the relative error. Indeed, let $(X_i)_{i\in\mathbb{N}}$ be a sequence of independent and identically distributed random variables with the same law as X. Then for any positive integer M,

$$\hat{p}_M = \frac{1}{M}\sum_{n=1}^{M} \mathbb{1}_{X_n \geq a} \tag{2}$$

is an unbiased estimator of p: $\mathbb{E}[\hat{p}_M] = p$. It is also well-known that its variance is given by $\mathrm{Var}(\hat{p}_M) = \frac{p(1-p)}{M}$ and therefore the relative error writes:

$$\frac{\sqrt{\mathrm{Var}(\hat{p}_M)}}{p} = \sqrt{\frac{1-p}{Mp}}. \tag{3}$$

Assume that the simulation of one random variable X_n requires a computational cost c_0. For a relative error of size ϵ, the cost of a direct Monte Carlo method is thus of the order

$$c_0 \frac{1-p}{\epsilon^2 p}. \tag{4}$$

which is prohibitive for small probabilities (say 10^{-9}).

Many algorithms devoted to the estimation of the probability of rare events have been proposed. Here we focus on the so-called Adaptive Multilevel Splitting (AMS) algorithm [4, 5].

Let us explain one way to understand this algorithm (see [2] for a more general presentation). The splitting strategy relies on the following remark. Let us introduce J intermediate levels: $a_0 = 0 < a_1 < \ldots < a_J = a$. The small probability p satisfies:

$$p = \prod_{j=1}^{J} p_j$$

where $p_j = \mathbb{P}(X > a_j | X > a_{j-1})$, with $1 \leq j \leq J-1$ and $p_J = \mathbb{P}(X \geq a | X > a_{J-1})$. In order to use this identity to build an estimator of p, one needs to (i) define appropriately the intermediate levels and (ii) find a way to sample according to the conditional distributions $\mathcal{L}(X|X > a_{j-1})$ to approximate each p_j using independent Monte Carlo procedures.

In this article, we will be interesting in the idealized case where we assume we have a way to draw independent samples according to the conditional distributions $\mathcal{L}(X|X > a_j)$, $j \in \{1, \ldots, J-1\}$. We will discuss at length this assumption below. It is then easy to check that for a given J, the variance is minimized when $p_1 = \ldots = p_J = p^{1/J}$, and that the associated variance is a decreasing function of J. It is thus natural to try to devise a method to find the levels a_j such that $p_1 = \ldots = p_J$. In AMS algorithms, the levels are defined in an adaptive and random way in order to satisfy (up to statistical fluctuations) the equality of the factors p_j. This is based on an interacting particle system approximating the quantiles $\mathbb{P}(X \geq a)$ using an empirical distribution. The version of the algorithm we study depends on two parameters: n and k. The first one denotes the total number of particles. The second one denotes the number of resampled particles at each iteration: they are those with the k lowest current levels (which mean that a sorting procedure is required). Thus, the levels are defined in such a way that $p_j = \left(1 - \frac{k}{n}\right)$ and the estimator of the probability p writes:

$$\hat{p}_{n,k} = C^{n,k}\left(1 - \frac{k}{n}\right)^{J^{n,k}}$$

where $J^{n,k}$ is the number of iterations required to reach the target level a, and $C^{n,k} \in \left[1 - \frac{k-1}{n}, 1\right]$ is a correction factor precisely defined below (see (10)). Notice that $C^{n,k} = 1$ if $k = 1$.

In all the following, we will make the following assumption:

Assumption 1.1. *X is a real-valued positive random variable which admits a continuous cumulative distribution function $t \mapsto \mathbb{P}(X \leq t)$.*

This ensures for example that (almost surely), there is no more than one particle at the same level, and thus that the resampling step in the algorithm is well defined. We will show in Section 2.3 below that this assumption can be relaxed: the continuity of the cumulative distribution function is actually only required on $[0, a)$, in which case the AMS algorithm still yields an estimate of $\mathbb{P}(X \geq a)$ (with a large inequality). From Section 3, we will always work under Assumption 1.1, and thus we will always use for simplicity strict rather than large inequalities on X (notice that under Assumption 1.1, $p = \mathbb{P}(X \geq a) = \mathbb{P}(X > a)$).

Our aim in this article is twofold. First, we show that for any values of n and k, the estimator $\hat{p}_{n,k}$ is an unbiased estimator of p (see Theorem 4.1):

$$\boxed{\mathbb{E}(\hat{p}_{n,k}) = p.} \tag{5}$$

Second, we are able to obtain an explicit asymptotic expression for the variance of the estimator $\hat{p}_{n,k}$ in the limit of large n, for fixed k and p, and thus for the relative error (see Proposition 5.2, Equation (36)):

$$\boxed{\frac{\sqrt{\mathrm{Var}(\hat{p}^{n,k})}}{p} = \frac{-\log p}{\sqrt{n}} \left(1 + \frac{(1 - \log(p))\,(k-1)}{2n} + \mathrm{o}\left(\frac{1}{n}\right)\right)^{1/2}.} \tag{6}$$

Thus, if we consider the cost associated to the Monte Carlo estimator based on M independent realizations of the algorithm, and if M is chosen in such a way that the relative error is of order ϵ, one ends up with the following asymptotic cost for the AMS algorithm (see Theorem 5.4, Equations (32) and (38)): for fixed k and p, in the limit of large n

$$\boxed{\frac{c_0 + c_1 \log(n)}{\epsilon^2} \left[\left((\log(p))^2 - \log(p)\right) + \frac{1}{n}\left(-\log(p)\,(k-1) + \frac{1}{2}\,(\log(p))^2 - \frac{1}{2}\,(\log(p))^3\right) + \mathrm{o}\left(\frac{1}{n}\right)\right]} \tag{7}$$

where c_0 denotes the cost for drawing one sample according to the conditional distributions $\mathcal{L}(X|X > x)$ (assumed to be independent of x, to simplify), and $c_1 \log(n)$ is the cost associated with the sorting procedures involved in the algorithm. Here again the Landau symbol o in the above depend non-uniformly on k and p. The two results (6) and (7) should be compared to the corresponding formulae for direct Monte Carlo simulation (3) and (4) above. From these results, we conclude that, in this asymptotic regime:

(i) the choice $k = 1$ is the optimal one in terms of variance and cost
(ii) AMS yields better result than direct Monte Carlo if

$$\left(1 + \frac{c_1}{c_0} \log(n)\right) \left((\log p)^2 - \log p\right) < \frac{1-p}{p}$$

which will be the case for sufficiently small probability p.

The assumption that we are able to sample according to the conditional distributions $\mathcal{L}(X|X > x)$ (idealized setting) is a severe limitation, from a practical viewpoint. Let us make three comments about this idealized setting. First, to the best of our knowledge, all the theoretical results which have been obtained so far in the literature [3, 4, 5, 7] rely on such an assumption. It is believed that the qualitative conclusions obtained under this assumption are still meaningful for the actual algorithm used in practice, where the conditional distributions $\mathcal{L}(X|X > a_{j-1})$ are only approximately sampled (using for example Metropolis Hastings procedures). In some sense, in the idealized setting, one studies the optimal performance one could reach with this algorithm. Second, we will describe in Section 2.2 situations where this assumption makes sense in practice: this is in particular the case in the 1d dynamic setting described in Section 2.2.2. Third, in a paper in preparation [2], we will actually show that it is possible to obtain an unbiased estimator of p using AMS in a very general setting. In other words, the idealized setting is crucial to obtain estimates on the variances and the costs of the algorithm, but not to prove the unbiasedness property for the estimator of the rare event probability.

Let us now review the results known from the literature on AMS. As mentioned above, the AMS algorithm has been introduced in [4, 5], where it is proven that their estimator $\hat{p}_{n,k}$ indeed converges almost surely to p (in the limit $n \to \infty$). A central limit theorem is also provided, in the asymptotic regime where $\frac{k}{n} = p_0$ is fixed. In [7], the authors consider the case $k = 1$, prove the unbiasedness of the estimator and analyze the variance. In [3], the authors analyze the large n limit, with fixed $\frac{k}{n} = p_0$. In summary, our results differ from what has been proven before in two ways: we prove that $\hat{p}_{n,k}$ is an unbiased estimator for any $k \geq 1$

(and not only $k = 1$) and we analyze the variance and the cost at a fixed k (in the limit $n \to \infty$) and show that in this regime, $k = 1$ is optimal.

In addition to the new results presented in this paper, we would like to stress that the techniques of proof we use seem to be original. The main idea is to consider (under Assumption 1.1) the family $(P(x))_{x\in[0,a]}$ of conditional probabilities:

$$P(x) = \mathbb{P}(X > a|X > x), \tag{8}$$

and to define associated estimators $\hat{p}_{n,k}(x)$ thanks to the AMS algorithm. We are then able to derive an *explicit functional equation on* $\mathbb{E}[\hat{p}_{n,k}(x)]$ *as a function of* x; which follows from the existence of an explicit expression for the distribution of order statistics of independent random variables. We can then check the unbiased property (see Theorem 4.1)

$$\mathbb{E}[\hat{p}_{n,k}(x)] = P(x).$$

Finally noting that $P(0) = p$, we get (5). To analyze the computational cost (Theorem 5.4), we follow a similar strategy, with several more technical steps. First, we derive functional equations for the variance and the mean number of iterations in the algorithm. In general we are not able to give explicit expressions to the solutions of these equations, and we require the following auxiliary arguments:

(i) We show how one can relate the general case to the so-called exponential case, when X has an exponential distribution with mean 1;
(ii) We then prove that the solutions of the functional equations are solutions of linear Ordinary Differential Equations of order k;
(iii) We finally get asymptotic results on the solutions to these ODEs in the limit $n \to +\infty$, with fixed values of p and k.

The paper is organized as follows. In Section 2, we introduce the AMS algorithm and discuss the sampling of the conditional distributions $\mathcal{L}(X|X > z)$. In Section 3, we show how to relate the case of a general distribution for X to the case when X is exponentially distributed. We then prove that the algorithm is well defined, in the sense that it terminates in a finite number of iterations (almost surely). In Section 4, we show one of our main result, namely Theorem 4.1 which states that the estimators $\hat{p}_{n,k}(x)$ are unbiased. Finally, in Section 5, we study the cost of the algorithm, with asymptotic expansions in the regime where p and k are fixed and n goes to infinity. The proofs of the results of Section 5 are postponed to Section 6.

2. The Adaptive Multilevel Splitting algorithm

After presenting the AMS algorithm in Section 2.1, we discuss the fundamental assumption that we know how to sample according to the conditional distributions $\mathcal{L}(X|X > z)$ in Section 2.2. We will in particular show that this assumption is actually practical at least in one setting: the one-dimensional dynamic setting presented in Section 2.2.2. Finally, we discuss Assumption 1.1 in Section 2.3 and show that it can be replaced by a less stringent hypothesis. This is particularly useful in the framework of the high-dimensional dynamic setting described in Section 2.2.3.

We would like to stress that Sections 2.2 and 2.3 discuss practical aspects of the implementation of AMS and can be skipped if the reader is only interested in the two main results about the unbiasedness and cost of AMS, presented in Sections 4 and 5.

2.1. Description of the algorithm.

We fix a total number n of particles, as well as $1 \in \{1, \dots, n-1\}$ the number of resampled particles at each iteration of the algorithm. In the sequel, when we consider a random variable X_i^j, the subscript i denotes the index in $\{1, \dots, n\}$ of a particle, and the superscript j denotes the iteration of the algorithm.

In the algorithm below and in the following, we use classical notations for k-th order statistics. For $Y = (Y_1, \dots, Y_n)$ an ensemble of independent and identically distributed (i.i.d.) real valued random variables with continuous cumulative distributions function, there exists almost surely a unique (random) permutation σ of $\{1, \dots, n\}$ such that $Y_{\sigma(1)} < \dots < Y_{\sigma(n)}$. For any $k \in \{1, \dots, n\}$, we then use the classical notation $Y_{(k)} = Y_{\sigma(k)}$ to denote the k-th order statistics of the sample Y.

For any $x \in [0, a]$, we define the Adaptive Multilevel Splitting algorithm as follows (in order to approximate $p = \mathbb{P}(X \geq a)$ one should take $x = 0$, but we consider the general case $x \in [0, a]$ for theoretical purposes).

Algorithm 2.1 (Adaptive Multilevel Splitting)**.**
Initialization: *Define $Z^0 = x$. Sample n i.i.d. realizations $X_1^0, \dots, X_n^0$, with the law $\mathcal{L}(X|X > x)$.*

Define $Z^1 = X^0_{(k)}$, the k-th order statistics of the sample $X^0 = (X_1^0, \dots, X_n^0)$, and σ^1 the (a.s.) unique associated permutation: $X^0_{\sigma^1(1)} < \dots < X^0_{\sigma^1(n)}$.

Set $j = 1$.
Iterations (on $j \geq 1$): *While $Z^j < a$:*

Conditionally on Z^j, sample k new independent [1] random variables $(\chi_1^j, \dots, \chi_k^j)$, according to the law $\mathcal{L}(X|X > Z^j)$.

Set

$$X_i^j = \begin{cases} \chi^j_{(\sigma^j)^{-1}(i)} & \text{if } (\sigma^j)^{-1}(i) \leq k \\ X_i^{j-1} & \text{if } (\sigma^j)^{-1}(i) > k. \end{cases}$$

In other words, the particle with index i is killed and resampled according to the law $\mathcal{L}(X|X > Z^j)$ if $X_i^{j-1} \leq Z^j$, and remains unchanged if $X_i^{j-1} > Z^j$. Notice that the condition $(\sigma^j)^{-1}(i) \leq k$ is equivalent to $i \in \{\sigma^j(1), \dots, \sigma^j(k)\}$.

Define $Z^{j+1} = X^j_{(k)}$, the k-th order statistics of the sample $X^j = (X_1^j, \dots, X_n^j)$, and σ^{j+1} the (a.s.) unique[2] associated permutation: $X^j_{\sigma^{j+1}(1)} < \dots < X^j_{\sigma^{j+1}(n)}$.

Finally increment $j \leftarrow j + 1$.
End of the algorithm: *Define $J^{n,k}(x) = j - 1$ as the (random) number of iterations. Notice that $J^{n,k}(x)$ is such that $Z^{J^{n,k}(x)} < a$ and $Z^{J^{n,k}(x)+1} \geq a$.*

The estimator of the probability p^x is defined by

$$\hat{p}^{n,k}(x) = C^{n,k}(x) \left(1 - \frac{k}{n}\right)^{J^{n,k}(x)}, \tag{9}$$

with

$$C^{n,k}(x) = \frac{1}{n} \mathrm{Card} \left\{ i;\, X_i^{J^{n,k}(x)} \geq a \right\}. \tag{10}$$

Notice that $C^{n,1}(x) = 1$. More generally, $C^{n,k}(x) \in \{\frac{n-k+1}{n}, \dots, \frac{n-k+k}{n}\}$.

Since we are interested in the algorithm starting at $x = 0$, we introduce the notation

$$\hat{p}^{n,k} = \hat{p}^{n,k}(0). \tag{11}$$

We finally stress that the computations of the sampled random variables $(X_i^0)_{1 \leq i \leq n}$ for the initialization and of the $(\chi_i^j)_{1 \leq i \leq k}$ for each iteration j can be made in parallel.

2.2. **Sampling from the conditional distributions $\mathcal{L}(X|X > z)$.** At each iteration of the algorithm, we need to sample k random variables according to conditional distributions $\mathcal{L}(X|X > z)$, with z taking values in the sequence $(Z^j)_{0 \leq j \leq J^{n,k}(x)}$. As explained above, we develop our theoretical analysis of the properties of the algorithm (bias, variance and computational cost) in the idealized situation where it is possible to sample according to these conditional distributions $\mathcal{L}(X|X > z)$ for any $z \in [0, a]$. From a practical point of view, this assumption is generally unrealistic. One possible situation where it is realistic is the dynamic setting presented in the present section (in contrast with the static setting).

2.2.1. *The static setting and the exponential case.* In a general framework, there is no simple way to sample the distributions $\mathcal{L}(X|X > z)$. In practice, this can be done thanks to a Metropolis-Hastings procedure, see for example [7]. Of course, this introduces a bias and correlations between the particles at each iteration (compared with the idealized algorithm studied in this paper). This bias and these correlations asymptotically vanish when the number of iterations in the Metropolis-Hastings procedure goes to infinity. The error analysis associated to this procedure is out of the scope of this paper.

[1] A precise mathematical statement is as follows. Let $(U_\ell^j)_{1 \leq \ell \leq k, j \in \mathbb{N}^*}$ be i.i.d. random variables, uniformly distributed on $(0, 1)$ and independent from all the other random variables. Then set $\chi_\ell^j = F(.; Z^j)^{-1}(U_\ell^j)$, where $F(.; x)^{-1}$ is the inverse distribution function associated with the (conditional) probability distribution $\mathcal{L}(X|X > x)$, see 19. We slightly abuse notation by using $\mathcal{L}(X|X > Z^j)$ rather than $\mathcal{L}(X|X > z)|_{z=Z^j}$.

[2] The uniqueness of the permutation σ^{j+1} is justified by Proposition 3.2.

There is a simple example where it is actually possible to sample the distributions $\mathcal{L}(X|X > z)$, namely if X is exponentially distributed. Indeed, if X has exponential law $\mathcal{E}(1)$ with mean 1, then the conditional distribution $\mathcal{L}(X|X > x) = \mathcal{L}(X + x)$ is a shifted exponential variable, for any $x > 0$. In the following, we will refer to this situation as *the exponential case*. Of course, this has no practical interest since in this case, $p = \mathbb{P}(X \geq a) = \exp(-a)$ is analytically known. However, this particular case plays a crucial role in the analysis hereafter, since as will be precisely explained in Section 3.3, the study of the general case can be reduced to the study of the exponential case after some change of variable. This trick was already used in the original papers [4, 5].

2.2.2. *Dynamic setting in dimension 1.* In the one-dimensional dynamic setting, X is defined as

$$X = \sup_{0 \leq t \leq \tau} Y_t$$

where $(Y_t)_{0 \leq t \leq \tau}$ is a strongly Markovian time-homogeneous random process with values in $\mathbb{R}$, and τ is a stopping time. In this setting, the conditional distribution $\mathcal{L}(X|X > x)$ is easily sampled: it is the law of $\sup_{0 \leq t \leq \tau} Y_t^x$, where Y_t^x denotes the stochastic process $(Y_t)_{t \geq 0}$ which is such that $Y_0 = x$.

Having in mind applications in molecular dynamics [6], a typical example is when $(Y_t)_{t \geq 0}$ satisfies a Stochastic Differential Equation

$$dY_t^x = f(Y_t^x)dt + \sqrt{2\beta^{-1}}dB_t, \quad Y_0^x = x,$$

with smooth drift coefficient f and inverse temperature $\beta > 0$. The stopping time is for example

$$\tau^x = \inf\left\{t \geq 0; Y_t^x < -\epsilon \text{ or } Y_t^x > 1 + \epsilon\right\},$$

for $x \in [0, 1]$, and for some $\epsilon > 0$ and one can then consider $X = \sup_{0 \leq t \leq \tau^0} Y_t^0$. Let us consider the target level $a = 1$. The AMS algorithm then yields an estimate of $\mathbb{P}(X \geq 1)$ the probability that the stochastic process starting from 0 reaches the level 1 before the level $-\epsilon$. Such computations are crucial to compute transition rates and study the so-called reactive paths in the context of molecular dynamics, see [6].

Notice that in practice, a discretization scheme must be employed, which makes the sampling of the conditional probabilities more complicated. Another point of view on the AMS algorithm is then required in order to prove the unbiasedness of the estimator of the probability p, see [2].

Remark 2.2. *The exponential case can be obtained from a dynamic setting. Indeed, consider the following stochastic process: a particle starts at a given position x, moves on the real line with speed $p = +1$ on the random interval $[0, \tau]$ where $\tau \sim \mathcal{E}(1)$ is exponentially distributed, and with speed $p = -1$ on the interval $(\tau, +\infty)$. More precisely,*

$$p_t^x = \begin{cases} +1 \text{ for } 0 \leq t \leq \tau \\ -1 \text{ for } t > \tau \end{cases} \qquad q_t^x = \begin{cases} x + t \text{ for } 0 \leq t \leq \tau \\ x + \tau - t \text{ for } t > \tau. \end{cases}$$

Notice that $(p_t^x, q_t^x)_{t \geq 0}$ is a Markov process such that $(q_t^x)_{t \geq 0}$ is continuous. Then for any initial condition x and any given threshold $a > x$ we have $\mathbb{P}(\sup_{t \geq 0} q_t^x > a) = \mathbb{P}(\tau > a - x) = \exp(x - a)$. In particular, $X = \sup_{t \geq 0} q_t^0 = \sup_{0 \leq t \leq \tau} q_t^0$ is an exponential random variable with parameter 1.

2.2.3. *Dynamic setting in higher dimension.* Let us consider again a strongly Markovian time-homogeneous stochastic process $(Y_t)_{t \geq 0}$, but with values in $\mathbb{R}^d$ for $d \geq 2$. In this case, the levels need to be defined using a (continuous) function $\xi : \mathbb{R}^d \to \mathbb{R}$, sometimes called a reaction coordinate in the context of molecular dynamics.

Let us focus for simplicity on the case when $(Y_t)_{t \geq 0}$ is solution of the stochastic differential equation (SDE):

$$dY_t^x = -\nabla V(Y_t^x)dt + \sqrt{2\beta^{-1}}dW_t, \quad Y_0^x = x, \tag{12}$$

with smooth potential V, inverse temperature $\beta > 0$ and $(W_t)_{t \geq 0}$ a d-dimensional Wiener process. Let us consider two disjoint closed subsets A and B of $\mathbb{R}^d$. Let us define the stopping time

$$\tau^x = \min(\tau_A^x, \tau_B^x)$$

where

$$\tau_A^x = \inf\left\{t \geq 0; Y_t^x \in A\right\} \text{ and } \tau_B^x = \inf\left\{t \geq 0; Y_t^x \in B\right\}. \tag{13}$$

Let us assume that the function ξ is such that

$$A = \{x;\, \xi(x) \leq 0\} \text{ and } B = \{x;\, \xi(x) \geq 1\}.$$

We then set, for a fixed initial condition $x_0 \in \mathbb{R}^d \setminus (A \cup B)$,

$$X = \sup_{0 \leq t \leq \tau} \xi(Y_t^{x_0}). \tag{14}$$

Let us set $a = 1$ as the target level. In this case, the probability $p = \mathbb{P}(X \geq 1) = \mathbb{P}(\tau_B^{x_0} < \tau_A^{x_0})$ is the probability that the path starting from x_0 reaches B before A. As explained above, this is a problem of interest in molecular dynamics for example, to study reactive paths and compute transition rates in high dimension, typically when A and B are metastable regions for $(Y_t)_{t \geq 0}$.

The problem to apply the AMS algorithm is again to sample according to the conditional distributions $\mathcal{L}(X|X > z)$. A natural idea is to use the following branching procedure in the resampling step at the j-th iteration: to build one of the new k trajectories, one of the $(n-k)$ remaining trajectories is chosen at random, copied up to the first time it reaches the level $\{x;\, \xi(x) = Z^j\}$ and then completed independently from the past up to the stopping time τ. The problem is that this yields in general a new trajectory which is correlated to the copied one through the initial condition on the level set $\{x;\, \xi(x) = Z^j\}$. Indeed, in general, given $x_1 \neq x_2$ such that $\xi(x_1) = \xi(x_2)$, the laws of $\sup_{0 \leq t \leq \tau} \xi(Y_t^{x_1})$ and $\sup_{0 \leq t \leq \tau} \xi(Y_t^{x_2})$ are not the same. As a consequence, it is unclear how to sample $\mathcal{L}(X|X > z)$, except if we would be able to build a function ξ such that the law of $\sup_{0 \leq t \leq \tau} \xi(Y_t^x)$ only depends on $\xi(x)$. This is actually the case if ξ is the so-called committor function associated to the dynamics (12) and the two sets A and B.

Definition 2.3. *Let A and B be two disjoint closed subsets in $\mathbb{R}^d$. The committor function ξ associated with the SDE* (12) *and the sets A and B is the unique solution of the following partial differential equation (PDE):*

$$\begin{cases} -\nabla V \cdot \nabla \xi + \beta^{-1} \Delta \xi = 0 & \text{in } \mathbb{R}^d \setminus (A \cup B), \\ \xi(x) = 0 & \text{for } x \in A, \\ \xi(x) = 1 & \text{for } x \in B. \end{cases} \tag{15}$$

Proposition 2.4. *Assume that ξ is the committor function, solution of* (15)*. For any $x \in \mathbb{R}^d \setminus (A \cup B)$, we set $X^x = \sup_{0 \leq t \leq \tau} \xi(Y_t^x)$. Notice that X^x is a random variable with values in $[\xi(x), 1]$.*

We have $\mathbb{P}(X^x > z) = \frac{\xi(x)}{z}$ for any $x \in \mathbb{R}^d \setminus (A \cup B)$ and $z \in [\xi(x), 1)$. In particular, if $\xi(x_1) = \xi(x_2)$, then we have the equality $\mathcal{L}(X^{x_1}) = \mathcal{L}(X^{x_2})$. Moreover, for any x_1, x_2 with $0 < \xi(x_1) \leq \xi(x_2) < 1$, $\mathcal{L}(X^{x_1}|X^{x_1} > \xi(x_2)) = \mathcal{L}(X^{x_2})$.

This previous Proposition (which is proven below) fully justifies the branching procedure described above to sample $\mathcal{L}(X^{x_0}|X^{x_0} > Z^j)$ at the j-th iteration of the algorithm: pick at random one of the $(n-k)$ remaining trajectories (say $(Y_t)_{t \geq 0}$), copy it up to the first time it reaches the level $\{x;\, \xi(x) = Z^j\}$ (let us denote Y_σ the first hitting point of this level) and then complete the trajectory, independently from the past. By the strong Markov property, this yields a new X sampled according to $\mathcal{L}(X^{Y_\sigma})$ which is indeed $\mathcal{L}(X^{x_0}|X^{x_0} > Z^j)$, since $\xi(Y_\sigma) = Z^j$.

Remark 2.5. *As already mentioned in the previous Section, in practice, the SDE* (12) *is discretized in time, say with a timestep Δt. Then, the branching procedure consists in copying the selected trajectory up to the first time index n such that $\xi(Y_{n\Delta t}) > z$, and then to complete it independently from the past. This introduces a difference compared to the continuous in time situation considered above, since doing so we do not sample according to the conditional distribution $\mathcal{L}(X|X > z)$. To treat this situation, one needs to resort to other techniques to analyze the algorithm, see* [2]*. In particular, one can show that the algorithm still yields an unbiased estimator of p, using very different techniques of proof than those presented in this paper. This approach is also useful to treat non-homogeneous in time Markov processes.*

At this stage, we can thus conclude that in the high-dimensional dynamic setting, if the committor function is known, the AMS algorithm can be practically implemented, and that it enters the framework of this paper. There are however two difficulties, that we will now discuss.

First, the random variable $X = \sup_{0 \leq t \leq \tau} \xi(Y_t^{x_0})$, where ξ is the committor function, does not satisfy Assumption 1.1: we have $p = \mathbb{P}(X \geq 1) = \mathbb{P}(X = 1) > 0$ (X takes values in $[0, 1]$) and therefore, the

cumulative distribution $t \mapsto \mathbb{P}(X \leq t)$ is not continuous at $t = 1$. More precisely, from Proposition 2.4, we have: $\forall t \in [0,1)$, $\mathbb{P}(X \leq t) = \left(1 - \frac{\xi(x_0)}{t}\right)_+$ and $\forall t \geq 1$, $\mathbb{P}(X \leq t) = 1$. This is actually not a problem, as explained in the next Section in a general setting: the continuity of the cumulative distribution function is only required over $[0,1)$ (or more generally over $[0,a)$ in the general case when the target level is a).

The second difficulty is that knowing the committor function is actually a very strong assumption. Computing ξ solution to (15) is actually impossible in practice since this is a high-dimensional PDE. Moreover, if ξ was known, then we would actually know the small probability we want to estimate since $p = \mathbb{P}(X \geq 1) = \xi(x_0)$. This is a consequence of the well-known probabilistic representation to solutions to (15):

Proposition 2.6. *Recall the definitions* (13) *of the stopping times τ_A^x and τ_B^x. Then, if ξ is the committor function associated to the SDE* (12) *and the sets A and B, then, for any $x \in \mathbb{R}^d$*

$$\xi(x) = \mathbb{P}(\tau_B^x < \tau_A^x). \tag{16}$$

Thus, this high-dimensional dynamic case with known committor function should also be considered as an idealized setting, which is only useful for theoretical purposes, in order to study the best performance we could expect for the AMS algorithm.

We end up this Section with a proof of Proposition 2.4.

Proof of Proposition 2.4: Let us consider ξ satisfying (15), Y_t^x solution to (12) and $X^x = \sup_{0 \leq t \leq \tau} \xi(Y_t^x)$. For any given $z \in (0,1)$, and any $x \in \mathbb{R}^d$, let us introduce

$$\tau_z^x = \inf\left\{t \geq 0; \xi(Y_t^x) \geq z\right\}.$$

One easily checks the identity

$$\mathbb{P}(X^x \geq z) = \mathbb{P}(X^x > z) = \mathbb{P}(\tau_z^x < \tau_A^x).$$

By continuity of ξ and of the trajectories of the stochastic process $(Y_t)_{t \geq 0}$, and by the strong Markov property at the stopping time τ_z^x, we get for any x and any $z \in (\xi(x), 1)$

$$\begin{aligned}
\xi(x) = \mathbb{P}(\tau_B^x < \tau_A^x) &= \mathbb{E}\left[\mathbb{1}_{\tau_z^x < \tau_A^x} \mathbb{1}_{\tau_B^x < \tau_A^x}\right] \\
&= \mathbb{E}\left[\mathbb{1}_{\tau_z^x < \tau_A^x} \mathbb{1}_{\tau_B^{Y^x_{\tau_z^x}} < \tau_A^{Y^x_{\tau_z^x}}}\right] \\
&= \mathbb{E}\left[\mathbb{1}_{\tau_z^x < \tau_A^x} \xi(Y^x_{\tau_z^x})\right] \\
&= \mathbb{P}(\tau_z^x < \tau_A^x) z.
\end{aligned}$$

This identity proves the first claim of the Proposition. Moreover, since the law of Y^x depends on x only through $\xi(x)$, we also more generally get $\mathcal{L}(X^{x_1} | X^{x_1} > \xi(x_2)) = \mathcal{L}(X^{x_2})$ for any x_1, x_2 with $0 < \xi(x_1) \leq \xi(x_2) < 1$,. □

2.3. **About Assumption 1.1.** In this section, we show that Assumption 1.1 is actually too stringent. If one assumes the following

Assumption 2.7. *X is a real-valued positive random variable such that $t \in [0,a) \mapsto \mathbb{P}(X \leq t)$ is continuous,*

then the Algorithm 2.1 is well defined, and all the results presented below hold. In particular, the estimator $\hat{p}_{n,k}$ is an unbiased estimator of

$$p = \mathbb{P}(X \geq a).$$

We notice that Assumption 2.7 is indeed more natural than Assumption 1.1 since the AMS algorithm only applies a resampling procedure with conditional distributions $\mathcal{L}(X | X > Z^j)$ to realizations such that $Z^j \in [0,a)$: this is why the continuity of the cumulative distribution function $t \mapsto \mathbb{P}(X \leq t)$ is actually only required over $[0,a)$.

The argument to show that one can recover the setting of Assumption 1.1 assuming only Assumption 2.7 is the following coupling argument. In Lemma 2.8 below, it is proven that there exists a random variable $\tilde{X}$ such that:

- $\tilde{X}$ satisfies Assumption 1.1;
- for any $z \in [0,a)$, $\tilde{X} \leq z$ is equivalent to $X \leq z$ and $\mathcal{L}(\tilde{X} | \tilde{X} > z) = \mathcal{L}(X | X > z)$;

- $\tilde{X} \geq a$ is equivalent to $X \geq a$ and thus, in particular, $\mathbb{P}(\tilde{X} \geq a) = \mathbb{P}(X \geq a) = p$, the probability to be estimated.

The last two properties show that running the AMS algorithm on $\tilde{X}$ is equivalent to running the AMS algorithm on X: the iterations, the stopping criterion and the estimator are the same. The theory developed in this paper (unbiased estimator, analysis of the cost and of the computational cost) is then applied to the algorithm applied to the auxiliary random variable $\tilde{X}$ instead of X, which is completely equivalent to the algorithm applied to X.

In all the following, for simplicity, we will always assume that Assumption 1.1 holds, keeping in mind that it can be relaxed to Assumption 2.7. Thus, inequalities which involve the random variable X can be changed from large to strict without modifying the associated events (almost surely).

We end this Section with a Lemma which defines the random variable $\tilde{X}$ as a function of X.

Lemma 2.8. *Let X be a random variable satisfying Assumption 2.7, and let us define*

$$\tilde{X} = X \mathbb{1}_{X<a} + \frac{a}{U} \mathbb{1}_{X \geq a},$$

where U is a random variable independent of X and uniformly distributed on $(0,1)$.

Then, (i) $\tilde{X}$ satisfies Assumption 1.1, (ii) for any $z \in [0,a)$, $\tilde{X} \leq z$ is equivalent to $X \leq z$, and the two laws $\mathcal{L}(\tilde{X}|\tilde{X} > z)$ and $\mathcal{L}(X|X > z)$ coincide on (z,a) and (iii) $\tilde{X} \geq a$ is equivalent to $X \geq a$.

Proof of Lemma 2.8: Since $a/U > a$, it is easy to check the items (ii) and (iii). Let us now consider the cumulative distribution of $\tilde{X}$.

For $t < a$, $\mathbb{P}(\tilde{X} \leq t) = \mathbb{P}(X \leq t)$ and thus $t \mapsto \mathbb{P}(\tilde{X} \leq t)$ is continuous for $t \in [0,a)$ by Assumption 2.7.

For $t > a$, $\mathbb{P}(\tilde{X} \leq t) = \mathbb{P}(X < a) + \mathbb{P}(U \geq a/t,\, X \geq a) = \mathbb{P}(X < a) + \mathbb{P}(X \geq a)\left(1 - \frac{a}{t}\right) = 1 - \frac{a}{t}\mathbb{P}(X \geq a)$ and thus $t \mapsto \mathbb{P}(\tilde{X} \leq t)$ is continuous for $t \in (a, +\infty)$.

Finally, with these expressions one easily checks left and right continuity of $t \mapsto \mathbb{P}(\tilde{X} \leq t)$ at a.

This concludes the proof of the fact that $\tilde{X}$ satisfies Assumption 1.1, and thus the proof of Lemma 2.8. □

3. Reduction to the exponential case and well-posedness of the algorithm

The aim of this Section (see Section 3.3) is to prove the well-posedness of the algorithm, namely the fact that $J^{n,k}(x)$ is almost surely finite, when the probability $p = P(0)$ to estimate is positive. The argument is based on the fact that the general case is related to the exponential case through a change of variable, see Section 3.2 (this will be instrumental in the rest of the paper). Section 3.1 first gives a few notation that will be useful below.

3.1. **Notation.**

3.1.1. *General notation.* We will use the following set of notations, associated to the random variable X satisfying Assumption 1.1.

We denote by F the cumulative distribution function of the random variable X: $F(t) = \mathbb{P}(X \leq t)$ for any $t \in \mathbb{R}$, and $F(0) = 0$. From Assumption 1.1, the function F is continuous. Notice that it ensures that if Y is an independent copy of X, then $\mathbb{P}(X = Y) = 0$, and thus, in the algorithm, there is only at most one sample at a given level.

We recall that our aim is to estimate the probability

$$p = \mathbb{P}(X \geq a) = \mathbb{P}(X > a), \tag{17}$$

given a threshold $a > 0$. More generally, we define for any $x \in [0,a]$

$$P(x) = \mathbb{P}(X > a | X > x), \tag{18}$$

so that we have $p = P(0)$. Notice that $P(a) = 1$.

For any $x \in [0,a]$, $\mathcal{L}(X|X > x)$ admits a cumulative distribution function $F(.;x)$, which satisfies: for any $y \in \mathbb{R}$,

$$F(y;x) = \frac{F(y) - F(x)}{1 - F(x)} \mathbb{1}_{y \geq x}. \tag{19}$$

While the variable x always denotes the parameter in the conditional distributions $\mathcal{L}(X|X>x)$, we use the variable y as a dummy variable in the associated densities and cumulative distribution functions.

By Assumption 1.1, $\mathbb{1}_{y\geq x}$ in the definition above can be replaced with $\mathbb{1}_{y>x}$. Notice that $F(y;0)=F(y)$. Moreover, with these notations, we have

$$P(x) = 1 - F(a;x). \tag{20}$$

An important tool in the following is the family of functions: for any $x\in[0,a]$ and any $y\in\mathbb{R}$

$$\begin{aligned} \Lambda(y;x) &= -\log(1-F(y;x)) \in [0,+\infty] \\ \Lambda(y) &= \Lambda(y;0) = -\log(1-F(y)). \end{aligned} \tag{21}$$

We remark the following identity: for $0\leq x\leq y\leq a$,

$$\Lambda(y;x) = \log(1-F(x)) - \log(1-F(y)) = \Lambda(y) - \Lambda(x).$$

3.1.2. *Specific notation when X admits a density f.* In some places, we will assume that X admits a density f with respect to the Lebesgue measure (which indeed implies Assumption 1.1). This assumption is in particular satisfied in the exponential case (namely when X is exponentially distributed), which we will consider in several arguments below to study the bias and the computational cost.

If X admits a density f, the law $\mathcal{L}(X|X>x)$ of X conditionally on $\{X>x\}$ also admits a density $f(.;x)$, which satisfies: for any $y\geq 0$,

$$f(y;x) = \frac{f(y)}{1-F(x)}\mathbb{1}_{y\geq x}. \tag{22}$$

Notice that $f(y)=f(y;0)$.

We finally introduce some notations about order statistics of samples of i.i.d. real random variables. If X admits a density f, then the k-th order statistics of an i.i.d. sample $(X_1,\dots,X_n)$ (distributed according to the law of X) admits a density $f_{n,k}$ which satisfies: for any $y\in\mathbb{R}$,

$$f_{n,k}(y) = k\binom{n}{k}F(y)^{k-1}f(y)\big(1-F(y)\big)^{n-k}.$$

The associated cumulative distribution function is $F_{n,k}(y)=\int_{-\infty}^{y} f_{n,k}(z)dz$. Likewise, we introduce notations for the density and the cumulative distribution function of order statistics for the law $\mathcal{L}(X|X>x)$: when $0\leq x\leq y\leq a$ we set

$$\begin{aligned} f_{n,k}(y;x) &= k\binom{n}{k}F(y;x)^{k-1}f(y;x)\big(1-F(y;x)\big)^{n-k}, \\ F_{n,k}(y;x) &= \int_x^y f_{n,k}(z;x)dz. \end{aligned} \tag{23}$$

3.2. **Reduction to the exponential case.** One of the key tool in the following is the reduction of the general case to the exponential case, thanks to the use of the function Λ, defined by (21). We recall that the exponential case refers to the case when X is distributed according to the exponential law with parameter 1, see Section 2.2.1. The basic remark is the following. Since by Assumption 1.1 the cumulative distribution function $F(.;x)$ is continuous, we have the following classical result, see for instance Proposition 2.2 in [1].

Lemma 3.1. *If $Y\sim\mathcal{L}(X|X>x)$, then $F(Y;x)$ is uniformly distributed on $(0,1)$, and thus $\Lambda(Y;x)$ has an exponential law with parameter 1.*

Let us first state a result on the algorithm without any stopping criterion.

Proposition 3.2. *Let us consider the sequence of random variables $((X_i^j)_{1\leq i\leq n}, Z^j)_{j\geq 0}$ generated by the AMS algorithm 2.1* without any stopping criterium.*Set $Y_i^j=\Lambda(X_i^j)$ and $S^j=\Lambda(Z^j)$. Then we have the following properties.*

(i) *For any $j\geq 0$, $(Y_i^j - S^j)_{1\leq i\leq n}$ is a family of i.i.d. exponentially distributed random variables, with parameter 1.*
(ii) *For any $j\geq 1$, $(Y_i^j - S^j)_{1\leq i\leq n}$ is independent of $(S^l - S^{l-1})_{1\leq l\leq j}$.*
(iii) *The sequence $(S^j - S^{j-1})_{j\geq 1}$ is i.i.d..*

As a consequence, in law, the sequence $((Y_i^j)_{1\le i\le n}, S^j)_{j\ge 0}$ *is equal to the sequence of random variables* $((X_i^j)_{1\le i\le n}, Z^j)_{j\ge 0}$ *obtained by the realization of the AMS algorithm* without any stopping criterion *in the exponential case, with initial condition* $Z^0 = \Lambda(x)$.

Proof of Proposition 3.2: The last assertion is a direct consequence of the three former items and of the fact that in the exponential case the function Λ is the identity mapping.

Item (iii) is a consequence of items (i) and (ii). It remains to prove jointly those two items, which is done by induction on j. When $j = 0$, the result follows from the way the algorithm is initialized: for each $1 \le i \le n$, $Y_i^0 = \Lambda(X_i^0)$ is exponentially distributed (see Lemma 3.1), and the independence property is clear. Assuming that the properties (i) and (ii) are satisfied for all $k \le j$, it is sufficient to prove that:

(i) $(Y_i^{j+1} - S^{j+1})_{1\le i\le n}$ are i.i.d. and exponentially distributed with mean 1;
(ii) $(Y_i^{j+1} - S^{j+1})_{1\le i\le n}$ is independent of $(S^l - S^{l-1})_{1\le l\le j+1}$.

For any positive real numbers $y_1, \dots, y_n$, and any $s^1, \dots, s^{j+1}$, this is equivalent to proving that

$$\begin{aligned} A :&= \mathbb{P}\left(Y_1^{j+1} - S^{j+1} > y_1, \dots, Y_n^{j+1} - S^{j+1} > y_n, S^1 - S^0 > s^1, \dots, S^{j+1} - S^j > s^{j+1}\right) \\ &= \exp\left(-(y_1 + \cdots + y_n)\right) \mathbb{P}(S^1 - S^0 > s^1, \dots, S^{j+1} - S^j > s^{j+1}). \end{aligned} \tag{24}$$

We want to decompose the probability with respect to the value of $\big(\sigma^{j+1}(i)\big)_{1\le i\le k}$. We recall that almost surely we have

$$Y^j_{\sigma^{j+1}(1)} < \ldots < Y^j_{\sigma^{j+1}(k)} = S^{j+1} < Y^j_{\sigma^{j+1}(k+1)} < \ldots < Y^j_{\sigma^{j+1}(n)}.$$

In fact, in order to preserve symmetry inside the groups of resampled and not-resampled particles, we decompose over the possible values for the random set $\big\{\sigma^{j+1}(1)\dots\sigma^{j+1}(k)\big\}$. We thus compute a sum over all partitions

$$\{1, \dots n\} = I_- \sqcup I_+,$$

such that $\mathrm{Card}\big(I_-\big) = k$.

$$\begin{aligned} A &= \mathbb{P}\Big(Y_1^{j+1} - S^{j+1} > y_1, \dots, Y_n^{j+1} - S^{j+1} > y_n, S^1 - S^0 > s^1, \dots, S^{j+1} - S^j > s^{j+1}\Big) \\ &= \sum_{\substack{I_-\subset\{1,\dots,n\} \\ \mathrm{Card}(I_-)=k}} \mathbb{P}\Big(Y_1^{j+1} - S^{j+1} > y_1, \dots, Y_n^{j+1} - S^{j+1} > y_n, \\ &\qquad\qquad S^1 - S^0 > s^1, \dots, S^{j+1} - S^j > s^{j+1}, \big\{\sigma^{j+1}(1), \dots, \sigma^{j+1}(k)\big\} = I_-\Big) \\ &= \sum_{\substack{I_-\subset\{1,\dots,n\} \\ \mathrm{Card}(I_-)=k}} \mathbb{P}\Big(\Big\{Y_i^{j+1} - S^{j+1} > y_i; i \in I_+\Big\}, S^1 - S^0 > s^1, \dots, S^{j+1} - S^j > s^{j+1}\Big) \\ &\qquad\qquad \prod_{i\in I_-} \mathbb{P}\Big(\Lambda(\chi^{j+1}) - \Lambda(Z^{j+1}) > y_i\Big). \end{aligned}$$

In the last line, we used the fact that by construction of the algorithm, on the event we consider, namely $\big\{\sigma^{j+1}(1), \dots, \sigma^{j+1}(k)\big\} = I_-$, we have the equality of random variables $Y_i^{j+1} = \Lambda\left(\chi^{j+1}_{(\sigma^{j+1})^{-1}(i)}\right), i \in I_-$. The latter are i.i.d. and independent of all the other random variables used at this stage of the algorithm. Moreover:

$$\prod_{i\in I_-} \mathbb{P}\Big(\Lambda(\chi^{j+1}) - \Lambda(Z^{j+1}) > y_i\Big) = \prod_{i\in I_-} \exp\big(-y_i\big).$$

Let us now introduce a notation: for $I_- \subset \{1, \cdots n\}$ we set

$$M^j_{I_-} := \max\left\{Y_i^j - S^j; i \in I_-\right\}.$$

Remark that on the event $\big\{\sigma^{j+1}(1), \dots, \sigma^{j+1}(k)\big\} = I_-$, almost surely, we have $Y_i^j = Y_i^{j+1}$ and $S^{j+1} - S^j = M^j_{I_-}$. Using the independence properties of the $Y_i^j - S^j, 1 \le i \le n$ (and thus of $M^j_{I_-}$) from the induction

hypothesis (ii), we obtain

$$\begin{aligned}
&\mathbb{P}\Big(\left\{Y^{j+1} - S^{j+1} > y_i; i \in I_+\right\}, S^1 - S^0 > s^1, \dots, S^{j+1} - S^j > s^{j+1}\Big) \\
&\quad = \mathbb{P}\Big(\Big\{Y_i^j - S^j - M_{I_-}^j > y_i; i \in I_+\Big\}, S^1 - S^0 > s^1, \dots, M_{I_-}^j > s^{j+1}\Big) \\
&\quad = \mathbb{P}\Big(\Big\{Y_i^j - S^j - M_{I_-}^j > y_i; i \in I_+\Big\}, M_{I_-}^j > s^{j+1}\Big)\mathbb{P}\Big(S^1 - S^0 > s^1, \dots, S^j - S^{j-1} > s^j\Big).
\end{aligned}$$

We can then integrate the $Y_i^j - S^j, i \in I_+$ using the induction hypothesis (i):

$$\begin{aligned}
&\mathbb{P}\Big(\Big\{Y_i^j - S^j - M_{I_-}^j > y_i; i \in I_+\Big\}, M_{I_-}^j > s^{j+1}\Big) \\
&\quad = \mathbb{E}\Big[\prod_{i\in I_+} \exp\Big(-y_i - M_{I_-}^j\Big)\mathbb{1}_{M_{I_-}^j > s^{j+1}}\Big] \\
&\quad = \prod_{i\in I_+} \exp(-y_i)\mathbb{E}\Big[\exp\Big(-(n-k)M_{I_-}^j\Big)\mathbb{1}_{M_{I_-}^j > s^{j+1}}\Big]
\end{aligned}$$

The proof is now complete since:

$$\begin{aligned}
A &= \mathbb{P}\Big(Y_1^{j+1} - S^{j+1} > y_1, \dots, Y_n^{j+1} - S^{j+1} > y_n, S^1 - S^0 > s^1, \dots, S^{j+1} - S^j > s^{j+1}\Big) \\
&= \sum_{\substack{I_-\subset\{1,\dots,n\} \\ \mathrm{Card}(I_-)=k}} \Big(\prod_{i\in I_-} \exp(-y_i)\Big)\Big(\prod_{i\notin I_-} \exp(-y_i)\Big)\mathbb{E}\Big[\exp\Big(-(n-k)M_{I_-}^j\Big)\mathbb{1}_{M_{I_-}^j > s^{j+1}}\Big] \\
&\qquad\qquad \mathbb{P}\Big(S^1 - S^0 > s^1, \dots, S^j - S^{j-1} > s^j\Big) \\
&= \exp\Big(\sum_{i=1}^n y_i\Big)\mathbb{P}\Big(S^1 - S^0 > s^1, \dots, S^j - S^{j-1} > s^j\Big) \sum_{\substack{I_-\subset\{1,\dots,n\} \\ \mathrm{Card}(I_-)=k}} \mathbb{E}\Big[\exp\Big(-(n-k)M_{I_-}^j\Big)\mathbb{1}_{M_{I_-}^j > s^{j+1}}\Big],
\end{aligned}$$

which proves (24).

In particular, taking $y_1 = \ldots = y_n = s^1 = \ldots = s^j = 0$, we see that

$$\sum_{\substack{I_-\subset\{1,\dots,n\} \\ \mathrm{Card}(I_-)=k}} \mathbb{E}\Big[\exp\Big(-(n-k)M_{I_-}^j\Big)\mathbb{1}_{M_{I_-}^j > s^{j+1}}\Big] = \mathbb{P}(S^{j+1} - S^j > s^{j+1}).$$

This concludes the proof of Proposition 3.2.

□

The next Lemma shows that the AMS algorithm applied to X with target level a and the AMS algorithm applied to $\Lambda(X)$ with target level $\Lambda(a)$ stop at the same iteration.

Lemma 3.3. *Let us consider the sequence of random variables $((X_i^j)_{1\le i\le n}, Z^j)_{j\ge 0}$ generated by the AMS algorithm 2.1* without any stopping criterium*, and set $Y_i^j = \Lambda(X_i^j)$ and $S^j = \Lambda(Z^j)$. For any $\alpha > 0$, almost surely, $\left\{S^j \ge \Lambda(\alpha)\right\} = \left\{Z^j \ge \alpha\right\}$.*

Proof of Lemma 3.3: : First, Λ is non-decreasing so that $\left\{Z^j \ge \alpha\right\} \subset \left\{S^j \ge \Lambda(\alpha)\right\}$. Moreover, one easily checks that $\left\{S^j \ge \Lambda(\alpha)\right\} \cap \left\{Z^j < \alpha\right\} \subset \left\{S^j = \Lambda(\alpha)\right\}$. From Proposition 3.2 we know that $S^j = S^0 + \sum_{\ell=1}^j \left(S^\ell - S^{\ell-1}\right)$ admits a density with respect to the Lebesgue measure, since the $S^\ell - S^{\ell-1}$ have the density of the k-th order statistics of independent exponentially distributed random variables with parameter 1. Therefore $\mathbb{P}\left(\left\{S^j > \Lambda(\alpha)\right\} \neq \left\{Z^j > \alpha\right\}\right) = 0$. □

A direct corollary of Proposition 3.2 and Lemma 3.3 is that the original problem reduces to the exponential case.

Corollary 3.4. *Consider the sequences of random variables $(X_i^j)_{1\le i\le n, 0\le j\le J^{n,k}(x)}$ and $(Z^j)_{0\le j\le J^{n,k}(x)+1}$ generated by the AMS algorithm 2.1. Set $Y_i^j = \Lambda(X_i^j)$ and $S^j = \Lambda(Z^j)$.*

Then, the sequences $(Y_i^j)_{1\le i\le n, 0\le j\le J^{n,k}(x)}$ *and* $(S^j)_{0\le j\le J^{n,k}(x)+1}$ *are equal in law to the sequences of random variables* $(X_i^j)_{1\le i\le J^{n,k}(\Lambda(x))}$ *and* $(Z^j)_{0\le j\le J^{n,k}(\Lambda(x))+1}$ *obtained by the realization of the AMS algorithm in the exponential case, with initial condition* $Z^0=\Lambda(x)$ *and target level* $\Lambda(a)$*.*

Finally, in the next Sections, we need the following result, which is a consequence of Proposition 3.2.

Corollary 3.5. *For any* $j\ge 0$*, conditionally on* Z^j*, the random variables* $(X_i^j)_{1\le i\le n}$ *are i.i.d. with law* $\mathcal{L}(X|X>Z^j)$*.*

Proof: Thanks to Proposition 3.2, we see that $(\Lambda(X_i^j)-\Lambda(Z^j))_{1\le i\le n}$ are i.i.d. and exponentially distributed with mean 1. Since $\Lambda(X_i^j)-\Lambda(Z^j)=\Lambda(X_i^j;Z^j)$, we observe that for any $x_1,\dots,x_n\in[Z^j,+\infty)$,

$$\begin{aligned}\mathbb{P}(X_1^j>x_1,\dots,X_n^j>x_n|Z^j)&=\mathbb{P}(\Lambda(X_1^j)>\Lambda(x_1),\dots,\Lambda(X_n^j)>\Lambda(x_n)|\Lambda(Z^j))\\&=\mathbb{P}(\Lambda(X_1^j)-\Lambda(Z^j)>\Lambda(x_1)-\Lambda(Z^j),\dots,\Lambda(X_n^j)-\Lambda(Z^j)>\Lambda(x_n)-\Lambda(Z^j)|\Lambda(Z^j))\\&=\exp(-(\Lambda(x_1)-\Lambda(Z^j))\dots\exp(-(\Lambda(x_n)-\Lambda(Z^j))\\&=(1-F(x_1;Z^j))\dots(1-F(x_n;Z^j)).\end{aligned}$$

This concludes the proof. □

3.3. **Well-posedness of the algorithm.** To ensure that the algorithm giving an estimator of the probability is well-defined, namely that it gives a result after a finite number of steps, we prove in this Section that $J^{n,k}(x)$ is almost surely finite, when the probability $p=P(0)$ is positive. The proof relies on the reduction to the exponential case explained in the previous Section.

Proposition 3.6. *Suppose* $p=P(0)>0$*. Then for any* $x\in[0,a]$*, we have* $P(x)=\mathbb{P}(X>a|X>x)>0$*, and for any integers* n *and* k *with* $1\le k<n$*, the number of iterations in the AMS algorithm is almost surely finite:* $J^{n,k}(x)<+\infty$ *a.s.*

Proof of Proposition 3.6: To prove this result, we consider the AMS Algorithm 2.1 without any stopping criterion (namely the condition $Z^j\ge a$). As a consequence, we define sequences of random variables with the iteration index $j\in\mathbb{N}$: we get $(X_i^j)_{j\ge0}$, for any $i\in\{1,\dots,n\}$ and $(Z^j)_{j\ge0}$. Proposition 3.6 is then equivalent to the following statement: almost surely, $\{j\ge0;Z^j\ge a\}\neq\emptyset$.

Thanks to Proposition 3.2, we write for any $j\ge0$

$$S^j=S^0+\sum_{\ell=1}^{j}R^\ell,$$

where $R^\ell=S^\ell-S^{\ell-1}$ are independent and identically distributed positive random variables, satisfying $\mathbb{E}R^\ell\in(0;+\infty)$. Indeed,

$$\mathbb{E}R^1\le\mathbb{E}\max_{1\le i\le n}Y_i^0\le\mathbb{E}\sum_{1\le i\le n}Y_i^0=n,$$

where we recall that $Y_i^0=\Lambda(X_i^0)$ are independent and exponentially distributed with parameter 1. To prove that $\mathbb{E}R^1>0$, we write

$$\mathbb{E}R^1\ge\mathbb{E}\min_{1\le i\le n}Y_i^0=1/n,$$

since it is easily checked that $\min_{1\le i\le n}Y_i^0$ has an exponential distribution, with mean $1/n$.

By the Strong Law of Large Numbers, when $j\to+\infty$, we have the almost sure convergence

$$\frac{S^j}{j}\to\mathbb{E}R^1,$$

which yields $S^j\to+\infty$, almost surely, when $j\to+\infty$. As a consequence, almost surely, there exists some $j\in\mathbb{N}$ such that $S^j\ge\Lambda(a)$. Using Lemma 3.3, this then implies that $Z^j\ge a$, and that $J^{n,k}(x)<+\infty$. □

In the case $k=1$, following the ideas in the proof of Proposition 3.6, one can easily identify the law of the number of iterations (see [7] for a similar result).

Proposition 3.7. *The random variable* $J^{n,1}(x)$ *has a Poisson distribution with mean* $-n\log(P(x))$*.*

Proof of Proposition 3.7: In the case $k = 1$, $R^1 = \min_{1\leq i\leq n} Y_i^0$ has an exponential distribution with mean $1/n$. We recall that $(R^\ell = S^\ell - S^{\ell-1})_{\ell\geq 0}$ is a sequence of independent and identically distributed random variables.

Let us introduce the Poisson process, with intensity n, associated with the sequence of independent and exponentially distributed increments $(R^j)_{\ell\geq 0}$:

$$P_t = \sum_{\ell=0}^{+\infty} \mathbb{1}_{S^\ell\leq t}.$$

Since $S^0 = \Lambda(x)$, we identify that

$$J^{n,1}(x) = P_{\Lambda(a)-\Lambda(x)} = P_{-\log(P(x))}.$$

The result follows since for any $t \geq 0$ P_t has a Poisson distribution with mean nt. □

4. The estimator $\hat{p}^{n,k}(x)$ is unbiased

Recall that X satisfies Assumption 1.1.

Let us fix a total number of replicas n, as well as $k \in \{1, \dots, n-1\}$ the number of killed and resampled replicas at each iteration. Given a level $a > 0$, we recall that the estimator of the conditional probability $P(x) = \mathbb{P}(X > a|X > x)$ for each value of $x \in [0, a]$ is $\hat{p}^{n,k}(x)$, defined by (9). We introduce the following notation:

$$p^{n,k}(x) = \mathbb{E}[\hat{p}^{n,k}(x)]. \tag{25}$$

Recall that we are specifically interested in estimating the probability $p = P(0)$, and that the introduction of $P(x)$ for x in the interval $[0, a]$ is a tool to prove that the estimator is unbiased. We write the result in its full generality, and then specify it to the estimation of p.

Theorem 4.1. *For any $k \in \{1, \dots, n-1\}$, for any $a > 0$, such that $p = \mathbb{P}(X > a) = P(0) > 0$, and any $x \in [0, a]$, $\hat{p}^{n,k}(x)$ is an unbiased estimator of the conditional probability $P(x)$:*

$$\mathbb{E}[\hat{p}^{n,k}(x)] = P(x). \tag{26}$$

In particular, when $x = 0$, we have $\mathbb{E}[\hat{p}^{n,k}] = p$.

From Section 3.2 (see Corollary 3.4), it is sufficient to prove the result in the exponential case. Indeed, let us assume that (26) holds in the exponential case, and let us consider the general case of a random variable X satisfying Assumption 1.1, then we have

$$\begin{aligned}\mathbb{E}[\hat{p}^{n,k}(x)] &= \mathbb{P}(\Lambda(X) > \Lambda(a)|\Lambda(X) > \Lambda(x)) \\ &= \exp(-\Lambda(a) + \Lambda(x)) \\ &= \mathbb{P}(X > a|X > x) = P(x).\end{aligned}$$

The first equality is a consequence of Corollary 3.4 and Theorem 4.1 in the exponential case. The third equality is a direct consequence of the definition (21) of Λ.

The aim of this section is thus to prove Theorem 4.1 in the exponential case. In all the following, we denote by $f(x) = \exp(-x)$ the density of X, and we will use the notation introduced in Section 3.1.2 above for the density of the k-th statistics. Actually, the proof given below is valid as soon as X has a density f: we do not use the specific form of the density, and this specific form would not make the argument easier.

The proof of this result is divided into two steps. First, we show that the function $x \mapsto p^{n,k}(x)$ is solution of a functional equation. Second, we show that the function $x \mapsto P(x)$ is its unique solution.

4.1. Derivation of the functional equation satisfied by $p^{n,k}$.

Proposition 4.2. *Let us assume that X admits a density f. Assume $p = P(0) > 0$. The function $x \in [0, a] \mapsto p^{n,k}(x)$ is solution of the following functional equation (with unknown q): for any $0 \leq x \leq a$*

$$q(x) = \int_x^a \left(1 - \frac{k}{n}\right) q(y) f_{n,k}(y; x)dy + \theta_p^{n,k}(x), \tag{27}$$

with

$$\theta_p^{n,k}(x) = \sum_{l=0}^{k-1} \frac{n-l}{n} \mathbb{P}\left(S(x)_{(l)}^n \leq a \leq S(x)_{(l+1)}^n \right), \tag{28}$$

where $(S(x)_l^n)_{1\leq l\leq n}$ *are independent and identically distributed with density* $f(.;x)$ *(see* (23)*), while for* $1 \leq l \leq k-1$, $S(x)_{(l)}^n$ *denotes the* l*-th order statistics of this* n*-sample. By convention, we set* $S(x)_{(0)}^n = x$.

Proof: The key idea is to decompose the expectation $\mathbb{E}[\hat{p}^{n,k}]$ according to the (random but almost surely finite) value of the number $J^{n,k}(x)$ of iterations. The function $\theta_p^{n,k}$ appears as the result of the algorithm when $J^{n,k}(x) = 0$, while the integral formulation corresponds to the case $J^{n,k}(x) > 0$. In the latter case, we then condition on the value of the first level $Z^1 = X_{(k)}^0$ and use Corollary 3.5.

More precisely, we have

$$p^{n,k}(x) = \mathbb{E}\left[\hat{p}^{n,k}(x)\right] = \mathbb{E}\left[\hat{p}^{n,k}(x)\mathbb{1}_{J^{n,k}(x)=0}\right] + \mathbb{E}\left[\hat{p}^{n,k}(x)\mathbb{1}_{J^{n,k}(x)>0}\right].$$

First, we have from (9) and (10)

$$\begin{aligned}\mathbb{E}\left[\hat{p}^{n,k}(x)\mathbb{1}_{J^{n,k}(x)=0}\right] &= \mathbb{E}\left[C^{n,k}(x)\mathbb{1}_{J^{n,k}(x)=0}\right] = \mathbb{E}\left[C^{n,k}(x)\mathbb{1}_{X_{(k)}^0>a}\right] \\ &= \sum_{l=0}^{k-1} \frac{n-l}{n} \mathbb{E}\left[\mathbb{1}_{X_{(l)}^0 \leq a < X_{(l+1)}^0}\right] = \theta_p^{n,k}(x).\end{aligned}$$

Second, using conditional expectation with respect to Z^1,

$$\begin{aligned}\mathbb{E}\left[\hat{p}^{n,k}(x)\mathbb{1}_{J^{n,k}(x)>0}\right] &= \mathbb{E}\left[\mathbb{E}\left[\left(1-\frac{k}{n}\right)\left(1-\frac{k}{n}\right)^{J^{n,k}(x)-1} C^{n,k}(x)|Z^1\right]\mathbb{1}_{Z^1\leq a}\right] \\ &= \mathbb{E}\left[\left(1-\frac{k}{n}\right)\mathbb{E}\left[\left(1-\frac{k}{n}\right)^{J^{n,k}(Z^1)} C^{n,k}(Z^1)|Z^1\right]\mathbb{1}_{Z^1\leq a}\right] \\ &= \left(1-\frac{k}{n}\right)\mathbb{E}\left[\mathbb{E}\left[\hat{p}^{n,k}(Z^1)|Z^1\right]\mathbb{1}_{Z^1\leq a}\right] \\ &= \left(1-\frac{k}{n}\right)\mathbb{E}\left[p^{n,k}(Z^1)\mathbb{1}_{Z^1\leq a}\right] \\ &= \left(1-\frac{k}{n}\right)\int_x^a p^{n,k}(y) f_{n,k}(y;x)dy,\end{aligned}$$

where on the event $\left\{Z^1 \leq a\right\}$ the equality $\mathbb{E}[(1-\frac{k}{n})^{J^{n,k}(x)-1}C^{n,k}(x)|Z^1] = \mathbb{E}[(1-\frac{k}{n})^{J^{n,k}(Z^1)}C^{n,k}(Z^1)|Z^1]$ is a consequence of Corollary 3.5, and of the fact that both $J^{n,k}(x)$ and $J^{n,k}(Z^1)$ are almost surely finite.

The intuition for this computation is that after the first step, if the algorithm does not stop, we just have to restart the algorithm from the level Z^1, and consider its associated estimator of the probability. The multiplication by $(1-\frac{k}{n})$ corresponds to the first iteration, which allows to go from level $Z^0 = x$ to level Z^1. □

We will need in the following a more explicit formula for $\theta_p^{n,k}(x)$.

Lemma 4.3. *We have for any* $x \in [0,a]$

$$\theta_p^{n,k}(x) = \left(1 - F(a;x)\right)\left(1 - F_{n-1,k}(a;x)\right). \tag{29}$$

Proof: We recall that $(S(x)_l^n)_{1\leq l\leq n}$ denotes a n-sample of i.i.d. random variables with law $\mathcal{L}(X|X>x)$, and that for any $l \in \{1,\ldots,n\}$ the random variable $S(x)_{(l)}^n$ denotes the l-th order statistics of this n-sample: almost surely we have

$$S(x)_{(1)}^n < \ldots < S(x)_{(n)}^n;$$

by convention we moreover have $S(x)_{(0)}^n = x$.

The proof is based on the partition of the n-sample into the $(n-1)$-sample $(S(x)_l^{n-1})_{1\leq l\leq n-1} := (S(x)_l^n)_{1\leq l\leq n-1}$ and the random variable $S(x)_n^n$. We express the probabilities appearing in the definition of

$\theta_p^{n,k}(x)$, using the cumulative distribution function of $S(x)_n^n$ and of the l-th order statistics $S(x)_{(l)}^{n-1}$ of the $(n-1)$-sample.

First, starting from (28), we write

$$\begin{aligned}
&\sum_{l=0}^{k-1} \frac{n-l}{n} \mathbb{P}\left(S(x)_{(l)}^n \le a \le S(x)_{(l+1)}^n \right) \\
&\qquad = \sum_{l=0}^{k-1} \frac{n-l}{n} \ (n!) \ \mathbb{P}\left(S(x)_1^n \le \ldots \le S(x)_l^n \le a \le S(x)_{l+1}^n \le \ldots \le S(x)_n^n \right) \\
&\qquad = \sum_{l=0}^{k-1} (n-l) \ ((n-1)!) \ \mathbb{P}\left(S(x)_1^n \le \ldots \le S(x)_l^n \le a \right) \mathbb{P}\left(a \le S(x)_{l+1}^n \le \ldots \le S(x)_n^n \right),
\end{aligned}$$

since the random variables $S(x)_l^n$ are independent, for $l \in \{1, \ldots, n\}$.

Now, for a fixed $l \in \{0, \ldots, k-1\}$, using the fact that $(S(x)_h^n)_{l+1 \le h \le n}$ are i.i.d. and changing the position of $S(x)_n^n$ in the ordered sample $S(x)_{l+1}^n \le \ldots \le S(x)_{n-1}^n$, we have for any $j \in \{l, \ldots, n-1\}$

$$\mathbb{P}(a \le S(x)_{l+1}^n \le \ldots \le S(x)_n^n) = \mathbb{P}(a \le S(x)_{l+1}^n \le \ldots \le S(x)_j^n \le S(x)_n^n \le S(x)_{j+1}^n \le \ldots \le S(x)_{n-1}^n),$$

with the convention that for $j = l$ the right-hand side above is $\mathbb{P}(a \le S(x)_n^n \le S(x)_{l+1}^n \le \ldots \le S(x)_{n-1}^n)$, while for $j = n-1$ it is $\mathbb{P}(a \le S(x)_{l+1}^n \le \ldots \le S(x)_{n-1}^n \le S(x)_n^n)$.

We obtain (since all the terms in the sum below are all the same)

$$\begin{aligned}
&(n-l)\mathbb{P}(a \le S(x)_{l+1}^n \le \ldots \le S(x)_n^n) \\
&\qquad = \sum_{j=l}^{n-1} \mathbb{P}(a \le S(x)_{l+1}^n \le \ldots \le S(x)_j^n \le S(x)_n^n \le S(x)_{j+1}^n \le \ldots \le S(x)_{n-1}^n) \\
&\qquad = \mathbb{P}(a \le S(x)_{l+1}^n \le \ldots \le S(x)_{n-1}^n, S(x)_n^n \ge a) \\
&\qquad = \mathbb{P}(a \le S(x)_{l+1}^n \le \ldots \le S(x)_{n-1}^n) \mathbb{P}(S(x)_n^n \ge a).
\end{aligned}$$

The last equality comes from independence, and the sum expresses the fact that there are $n-l$ positions to insert $S(x)_n^n$ in the increasing sequence $S(x)_{l+1}^n \le \ldots \le S(x)_{n-1}^n$.

Thus

$$\begin{aligned}
&\sum_{l=0}^{k-1} \frac{n-l}{n} \mathbb{P}\left(S(x)_{(l)}^n \le a \le S(x)_{(l+1)}^n \right) \\
&\qquad = \sum_{l=0}^{k-1} (n-1)! \, \mathbb{P}\left(S(x)_1^n \le \ldots \le S(x)_l^n \le a) \mathbb{P}(a \le S(x)_{l+1}^n \le \ldots \le S(x)_{n-1}^n) \mathbb{P}(S(x)_n^n \ge a \right) \\
&\qquad = \mathbb{P}(S(x)_n^n \ge a) \sum_{l=0}^{k-1} (n-1)! \, \mathbb{P}\left(S(x)_1^{n-1} \le \ldots \le S(x)_l^{n-1} \le a \le S(x)_{l+1}^{n-1} \le \ldots \le S(x)_{n-1}^{n-1} \right) \\
&\qquad = \mathbb{P}(S(x)_n^n \ge a) \sum_{l=0}^{k-1} \mathbb{P}\left(S(x)_{(l)}^{n-1} \le a \le S(x)_{(l+1)}^{n-1} \right) \\
&\qquad = \mathbb{P}(S(x)_n^n \ge a) \mathbb{P}(S(x)_{(k)}^{n-1} \ge a) \\
&\qquad = \left(1 - F(a;x)\right) \left(1 - F_{n-1,k}(a;x)\right).
\end{aligned}$$

This concludes the proof of Lemma 4.3.

Notice that we have proved a stronger statement: for $l \in \{0, \ldots, k-1\}$

$$\frac{n-l}{n} \mathbb{P}\left(S(x)_{(l)}^n \le a \le S(x)_{(l+1)}^n \right) = \mathbb{P}(S(x)_n^n \ge a) \mathbb{P}\left(S(x)_{(l)}^{n-1} \le a \le S(x)_{(l+1)}^{n-1} \right). \tag{30}$$

□

4.2. **Proof of Theorem 4.1.** Let us first state a uniqueness result.

Lemma 4.4. *The functional equation* (27) *admits at most one solution* $p : [0,a] \to \mathbb{R}^+$ *in* $L^\infty([0,a])$.

Proof: Let $p_1, p_2 : [0,a] \to \mathbb{R}^+$ be two bounded solutions. Then, we have for any $x \in [0,a]$

$$\begin{aligned}|p_1(x) - p_2(x)| &\leq \left(1 - \frac{k}{n}\right) \int_x^a |p_1(y) - p_2(y)| f_{n,k}(y;x) dy \\ &\leq \left(1 - \frac{k}{n}\right) \|p_1 - p_2\|_\infty \int_x^a f_{n,k}(y;x) dy \\ &\leq \left(1 - \frac{k}{n}\right) \|p_1 - p_2\|_\infty,\end{aligned}$$

which shows that $p_1 = p_2$, since $k/n > 0$. □

Notice that both functions $p^{n,k}$ and P take values in $[0,1]$, and are therefore bounded. Thanks to Proposition 4.2, $p^{n,k}$ satisfies (27), and Theorem 4.1 is thus a direct consequence of Lemma 4.4 if we prove that P is also solution of this functional equation.

Proof of Theorem 4.1:

The proof consists in proving that $x \mapsto P(x) = 1 - F(a;x)$ is solution of (27). For this we have to compute for $x \in [0,a]$,

$$\begin{aligned}\int_x^a \left(1 - \frac{k}{n}\right) P(y) f_{n,k}(y;x) dy &= \int_x^a (1 - F(a;y)) \frac{(n-k)k}{n} \binom{n}{k} F(y;x)^{k-1} f(y;x) (1 - F(y;x))^{n-k} dy \\ &= \int_x^a (1 - F(a;y)) (1 - F(y;x)) k \binom{n-1}{k} F(y;x)^{k-1} f(y;x) (1 - F(y;x))^{n-k-1} dy \\ &= \int_x^a (1 - F(a;y)) (1 - F(y;x)) f_{n-1,k}(y;x) dy \\ &= (1 - F(a;x)) \int_x^a f_{n-1,k}(y;x) dy \\ &= (1 - F(a;x)) F_{n-1,k}(a;x),\end{aligned}$$

thanks to the definitions (23) of $f_{n,k}$ and $f_{n-1,k}$, and the relation

$$(1 - F(a;y)) (1 - F(y;x)) = 1 - F(a;x)$$

for any $x \leq y \leq a$, obtained from (19).

We then conclude by checking the following identity, which is a consequence of Lemma 4.3.

$$\begin{aligned}\int_x^a \left(1 - \frac{k}{n}\right) P(y) f_{n,k}(y;x) dy + \theta_p^{n,k}(x) &= (1 - F(a;x)) F_{n-1,k}(a;x) + (1 - F(a;x)) (1 - F_{n-1,k}(a;x)) \\ &= 1 - F(a;x) = P(x)\end{aligned}$$

which concludes the proof. □

5. Variance and computational cost: results

In this Section, we introduce a notion of cost for the algorithm (related to the variance of the estimator and to the expected number of iterations), which allows to study the influence of the parameters n and k. We then give asymptotic expansions of the variance, the expected number of iterations, and the cost, when n tends to $+\infty$, for fixed values of k and of the probability p, and give interpretations of the results, compared to a direct Monte Carlo estimate as presented in the Introduction.

All these results are given under Assumption 1.1. The proofs are then given in Section 6.

5.1. **Definition of the cost.** In the following, we denote by c_0 the cost corresponding to the simulation of one random variable sampled according to the law $\mathcal{L}(X|X > x)$, for any $x \in [0,a]$. We assume that this cost does not depend on x.

We consider the Monte Carlo approximation of p using M independent realizations of the AMS algorithm 2.1. The associated estimator is

$$\overline{p}_M^{n,k}(x) = \frac{1}{M}\sum_{m=1}^{M} \hat{p}_m^{n,k}(x), \tag{31}$$

where $\hat{p}_m^{n,k}(x)$ is the AMS estimator for the m-th independent realization of the algorithm. Following the reasoning used in the introduction on the direct Monte Carlo estimator, for a given tolerance error ϵ, we want the relative error

$$\frac{\left(\mathrm{Var}(\overline{p}_M^{n,k}(x))\right)^{1/2}}{P(x)}$$

to be less than ϵ, i.e.

$$\frac{1}{M}\mathrm{Var}(\hat{p}^{n,k}(x)) \leq \epsilon^2 P(x)^2.$$

We thus have to choose $M = \frac{\mathrm{Var}(\hat{p}^{n,k}(x))}{\epsilon^2 P(x)^2}$.

For each realization m of the algorithm, $J_m^{n,k}(x)$ iterations are necessary, so that $k\, J_m^{n,k}(x) + n$ random variables are sampled: n at the initial step, and then k new ones at each iteration. This gives a cost $c_0\left(k\, J_m^{n,k}(x) + n\right)$, where c_0 is the computational cost of the sampling of one random variable distributed according to $\mathcal{L}(X|X > x)$. Moreover, at the first iteration, one needs to sort the random variables $(X_1^0, \ldots, X_n^0)$ (with an associated cost $c_1 n \log n$) and at each iteration, one has to insert the k new sampled particles into the already sorted $(n-k)$ remaining particles (with an associated cost $c_1 k \log n$). The sorting procedures are thus associated with a cost $c_1(\log n)\left(k\, J_m^{n,k}(x) + n\right)$. The total cost is thus

$$\sum_{m=1}^{M} (c_0 + c_1 \log n)\left(k\, J_m^{n,k}(x) + n\right)$$

and by an application of the Law of Large Numbers, when M is large, it is legitimate to consider that the cost to obtain a relative error of size ϵ is thus

$$(c_0 + c_1 \log n)\frac{\mathbf{C}^{n,k}(x)}{\epsilon^2} \tag{32}$$

where

$$\mathbf{C}^{n,k}(x) = \frac{\mathrm{Var}(\hat{p}^{n,k}(x))\left(k\,\mathbb{E}[J^{n,k}(x)] + n\right)}{P(x)^2}.$$

This is consistent with the standard definition of the efficiency of a Monte Carlo procedure as "inversely proportional to the product of the sampling variance and the amount of labour expended in obtaining this estimate", see [8, Section 2.5]. This should be compared with the cost of a direct Monte Carlo computation, which we recall (see (4)):

$$\frac{(1-P(x))}{\epsilon^2 P(x)} c_0. \tag{33}$$

Remark 5.1. *If we add the possibility of using $N \geq 1$ processors to sample in parallel the required random variables at each iteration (assuming for simplicity that k/N is an integer) then the cost is divided by N, and is thus $(c_0 + c_1 \log n)\frac{\mathbf{C}^{n,k}(x)}{\epsilon^2 N}$. Notice that this resulting cost is the same as if we run in parallel N independent realizations of $\hat{p}_m^{n,k}(x)$ to compute the estimator $\overline{p}_M^{n,k}(x)$ (assuming for simplicity that M/N is an integer). Both these parallelization strategies have the same effect on the cost in the setting of this article.*

Let us set a few notations: for $x \in [0, a]$,

$$v^{n,k}(x) = \mathbb{E}\left[(\hat{p}^{n,k}(x))^2\right] \text{ and } T^{n,k}(x) = \mathbb{E}\left[J^{n,k}(x)\right] + 1. \tag{34}$$

Notice that $T^{n,k}(x)$ is the expected number of steps in the algorithm (the initialization plus $J^{n,k}(x)$ iterations). Using this notation, we have

$$\mathbf{C}^{n,k}(x) = \frac{v^{n,k}(x) - P(x)^2}{P(x)^2}\left(k\, T^{n,k}(x) + n - k\right). \tag{35}$$

5.2. **Asymptotic expansions of the variance and of the computational cost.** We divide the results into three parts. We first study the variance, and then the average number of iterations. Finally, we combine the results to get the cost. We do not have explicit expressions for each value of k and n, but we get informative asymptotic results when $n \to +\infty$.

We assume that $p = P(0) > 0$, and consider $x \in [0, a)$, such that $P(x) < 1$. Note that if $P(x) = 1$, then almost surely $p^{n,k}(x) = 1$, $\mathrm{Var}(\hat{p}^{n,k}(a)) = 0$ and $T^{n,k}(a) = 1$, so that no asymptotic expansion is necessary.

Proposition 5.2. *For any fixed k and any $0 \leq x < a$ with $P(x) < 1$, when $n \to +\infty$ we have*

$$\mathrm{Var}(\hat{p}^{n,k}(x)) = \frac{P(x)^2}{n}\left(-\log(P(x)) + \frac{\left[\left(\left(\log(P(x))\right)^2 - \log(P(x))\right](k-1)\right.}{2n} + \mathrm{o}\left(\frac{1}{n}\right)\right). \tag{36}$$

Proposition 5.3. *For any fixed k and $0 \leq x < a$ with $P(x) < 1$, when $n \to +\infty$ we have*

$$T^{n,k}(x) = n\left(-\log(P(x))\left[\frac{1}{k} - \frac{k-1}{2kn}\right] + \frac{3k-1}{2kn} + \mathrm{o}\left(\frac{1}{n}\right)\right). \tag{37}$$

Finally, we have the following result on the cost:

Theorem 5.4. *For any fixed k and $0 \leq x < a$ with $P(x) < 1$, when $n \to +\infty$ we have*

$$\begin{aligned}\mathbf{C}^{n,k}(x) = &\left[\left(\log(P(x))\right)^2 - \log(P(x))\right] \\ &+ \frac{1}{n}\left(-\log(P(x))\left[k-1\right] + \frac{1}{2}\left(\log(P(x))\right)^2 - \frac{1}{2}\left(\log(P(x))\right)^3\right) + \mathrm{o}\left(\frac{1}{n}\right).\end{aligned} \tag{38}$$

The proof of Theorem 5.4 from the Propositions 5.2 and 5.3 is straightforward using (35). The proof of the two Propositions is long and technical, and is postponed to Section 6.

Let us also state an immediate corollary of Theorem 4.1 and of Proposition 5.2.

Corollary 5.5. *For any fixed k and $0 \leq x < a$ with $P(x) < 1$, when $n \to +\infty$, we have the following convergence in probability:*

$$\hat{p}^{n,k}(x) \to P(x).$$

Remark 5.6. *The case $k = 1$ is much simpler than the general case $k > 1$, and allows for direct computations. As seen in Section 3.3, for $n \geq 2$ and $0 \leq x \leq a$, $J^{n,1}(x)$ has a Poisson distribution with parameter $-n\log(P(x))$. We can then easily check the unbiased property $p^{n,1}(x) = P(x)$, and compute $T^{n,1}(x) = -n\log(P(x)) + 1$ and $Var(\hat{p}^{n,k}(x)) = P(x)^2(P(x)^{-1/n} - 1)$. In particular, no asymptotic expansions are required in order to understand the behavior of the computational cost.*

Notice that in the case $k = 1$, $\mathbb{E}\left[J^{n,1}(x)\right] = -n\log(P(x))$, and thus $\frac{J^{n,1}(x)}{n}$ is an unbiased estimator of $-\log(P(x))$, with variance $\frac{-\log(P(x))}{n}$. There is no such statement when $k > 2$: from Proposition 5.3, we have the limit $\mathbb{E}\left[\frac{J^{n,k}}{n}\right] \to \frac{-\log(P(x))}{k}$, but the first-order term in $1/n$ is equal to $\frac{k-1}{2k}(1 + \log(P(x)))$ and is therefore not zero, except if $k = 1$ or $P(x) = \exp(-1)$ - in which case one should compute an higher-order expansion to prove that there is a bias.

Let us now recall the main two practical consequences of Theorem 5.4, that we already stressed in the introduction. First, in view of (33) and (32)-(38), the AMS algorithm is more efficient than a direct Monte Carlo procedure to estimate $p = P(0)$ if

$$\frac{1-p}{p}c_0 > (c_0 + c_1\log(n))\left((\log p)^2 - \log p\right)$$

which is always true for sufficiently small p.

Second, from Theorem 5.4, we observe that all choices of k give the same leading order term for the cost. But looking at the term of order $\frac{1}{n}$, we see that the optimal choice is $k = 1$. This conclusion can also be deduced from the asymptotic expansion on the variance given in Proposition 5.2.

6. Proof of the variance and computational cost estimates

This Section is devoted to the proof of Propositions 5.2 and 5.3, which together yield the cost estimate of Theorem 5.4. The main steps are the following:

- In Section 6.1, we first show that we can reduce the analysis to the exponential case, by the change of variable using the function Λ, as explained in Section 3.2.
- In Section 6.2, we then derive functional equations on the two functions $v^{n,k}$ and $T^{n,k}$ (defined by (34)). These functional equations can actually be obtained not only in the exponential case, but for any X which admits a density with respect to the Lebesgue measure.
- In Section 6.3, we prove that, in the exponential case, the functional equations on $v^{n,k}$ and $T^{n,k}$ are equivalent to linear ordinary differential equations of order k.
- Finally, we compute asymptotic expansions of $v^{n,k}$ (in Section 6.4) and $T^{n,k}$ (in Section 6.5) in the large n limit, for fixed k and p.

The main simplification provided by the exponential case is that the functional equations can be recast as *linear* differential equations (see Remark 6.5 below).

6.1. **Reduction to the exponential case.** We have seen in Section 3.2 (see Corollary 3.4) that the estimator $\hat{p}^{n,k}(x)$ obtained with the AMS algorithm applied to a general random variable X (satisfying Assumption 1.1) with initial condition x and target level a is exactly the same in law as the estimator $\hat{p}^{n,k}(\Lambda(x))$ that is obtained with the AMS algorithm applied to an exponentially distributed random variable X, with initial condition $\Lambda(x)$ and target level $\Lambda(a)$.

It is therefore sufficient to prove the Propositions 5.2 and 5.3 in the exponential case. Indeed, if we obtain in the exponential case, for an initial condition x and a target level $a \geq x$:

$$\mathrm{Var}(\hat{p}^{n,k}(x)) = \frac{\exp(2(x-a))}{n}\left((a-x) + \frac{\left[(a-x)^2 + (a-x)\right](k-1)}{2n} + \mathrm{o}\left(\frac{1}{n}\right)\right), \tag{39}$$

$$T^{n,k}(x) = n\left((a-x)\left[\frac{1}{k} - \frac{k-1}{2kn}\right] + \frac{3k-1}{2kn} + \mathrm{o}\left(\frac{1}{n}\right)\right), \tag{40}$$

then the general case is easily obtained by replacing x by $\Lambda(x)$ and a by $\Lambda(a)$, since $\Lambda(x) - \Lambda(a) = -\Lambda(a;x) = \log(1 - F(a;x)) = \log(P(x))$.

6.2. **Functional equations satisfied by $v^{n,k}$ and $T^{n,k}$.** We now write functional equations satisfied by $v^{n,k}(x)$ and $T^{n,k}(x)$ defined by (34). Even though we will only need these functional equations in the exponential case as explained above, we derive these functional equations in a more general setting, namely when X admits a density f with respect to the Lebesgue measure. We refer to Section 3.1.2 for relevant notation associated to this setting. Notice that the derivations of these functional equations are very similar to the derivation of the functional equation on $p^{n,k}$ in the proof of Proposition 4.2.

Proposition 6.1. *Assume $P(0) > 0$. The function $x \mapsto v^{n,k}(x)$ is solution of the following functional equation (with unknown w): for any $0 \leq x \leq a$*

$$w(x) = \int_x^a \left(1 - \frac{k}{n}\right) w(y) f_{n,k}(y;x)dy + \theta_v^{n,k}(x) \tag{41}$$

with

$$\theta_v^{n,k}(x) = \sum_{l=0}^{k-1} \frac{(n-l)^2}{n^2} \mathbb{P}\left(S(x)_{(l)}^n \leq a \leq S(x)_{(l+1)}^n\right). \tag{42}$$

where $(S(x)_l^n)_{1\leq l\leq n}$ are independent and identically distributed with density $f(x,.)$, and $S(x)_{(l)}^n$ denotes the l-th order statistics of this n-sample. By convention, $S(x)_{(0)}^n = x$.

Similarly, we can derive a functional equation satisfied by $T^{n,k}$.

Proposition 6.2. *Assume $P(0) > 0$. The function $x \mapsto T^{n,k}(x)$ is solution of the following functional equation (with unknown T): for any $0 \leq x \leq a$*

$$T(x) = \int_x^a T(y) f_{n,k}(y;x)dy + 1. \tag{43}$$

We do not give the details of the proofs of these two results. The proof of Proposition 6.1 follows exactly the same lines as the proof of Proposition 4.2. For Proposition 6.2, using the same arguments, one obtains:

$$\begin{aligned} T^{n,k}(x) &= \int_x^a (1 + T^{n,k}(y)) f_{n,k}(y;x)dy + \int_a^{+\infty} f_{n,k}(y;x)dy \\ &= \int_x^a T^{n,k}(y) f_{n,k}(y;x)dy + 1 \end{aligned}$$

which is indeed (43).

Similarly to Lemma 4.3 on $\theta_p^{n,k}$, we have a more explicit formula for $\theta_v^{n,k}$ that will be useful below.

Lemma 6.3. *Assume $1 \leq k \leq n-2$. We have for any $x \in [0,a]$*

$$\theta_v^{n,k}(x) = \frac{1}{n}\left(1 - F(a;x)\right)\left(1 - F_{n-1,k}(a;x)\right) + \left(1 - \frac{1}{n}\right)\left(1 - F(a;x)\right)^2\left(1 - F_{n-2,k}(a;x)\right). \tag{44}$$

Proof of Lemma 6.3: The notations are the following: $S(x)_1^n, \ldots, S(x)_n^n$ are n independent random variables distributed according to $\mathcal{L}(X|X > x)$, $S(x)_{(l)}^{n-1}$ is the l-th order statistic of the random variables $S(x)_1^{n-1} = S(x)_1^n, \ldots, S(x)_{n-1}^{n-1} = S(x)_{n-1}^n$, and $S(x)_{(l)}^{n-2}$ is the l-th order statistic of the random variables $S(x)_1^{n-2} = S(x)_1^n, \ldots, S(x)_{n-2}^{n-2} = S(x)_{n-2}^n$. Here again $S(x)_{(0)}^n = S(x)_{(0)}^{n-1} = x$.

From (42), and using twice the equality (30) obtained in the proof of Lemma 4.3, we have

$$\begin{aligned} \theta_v^{n,k}(x) &= \sum_{l=0}^{k-1} \frac{(n-l)^2}{n^2} \mathbb{P}\left(S(x)_{(l)}^n \leq a \leq S(x)_{(l+1)}^n\right) \\ &= \mathbb{P}\left(S(x)_n^n \geq a\right) \sum_{l=0}^{k-1} \frac{(n-l)}{n} \mathbb{P}\left(S(x)_{(l)}^{n-1} \leq a \leq S(x)_{(l+1)}^{n-1}\right) \\ &= \mathbb{P}\left(S(x)_n^n \geq a\right) \frac{n-1}{n} \sum_{l=0}^{k-1} \frac{(n-1-l)}{n-1} \mathbb{P}\left(S(x)_{(l)}^{n-1} \leq a \leq S(x)_{(l+1)}^{n-1}\right) \\ &\quad + \mathbb{P}\left(S(x)_n^n \geq a\right) \sum_{l=0}^{k-1} \frac{1}{n} \mathbb{P}\left(S(x)_{(l)}^{n-1} \leq a \leq S(x)_{(l+1)}^{n-1}\right) \\ &= \left(1 - \frac{1}{n}\right) \mathbb{P}\left(S(x)_n^n \geq a\right) \mathbb{P}\left(S(x)_{n-1}^{n-1} \geq a\right) \mathbb{P}\left(S(x)_{(k)}^{n-2} \geq a\right) \\ &\quad + \frac{1}{n} \mathbb{P}\left(S(x)_n^n \geq a\right) \mathbb{P}\left(S(x)_{(k)}^{n-1} \geq a\right), \end{aligned}$$

writing that $\frac{n-l}{n} = \frac{n-1-l}{n-1}\frac{n-1}{n} + \frac{1}{n}$. □

As mentioned above, the functional equations (41)-(43) and the equation (44) on $\theta_v^{n,k}$ are valid for any X with a density f. However, we are only able to exploit them in the exponential case. From now on, we thus only consider the exponential case: $X \sim \mathcal{E}(1)$, $f(x) = \exp(-x)\mathbb{1}_{x \geq 0}$ and $F(x) = (1 - \exp(-x))\mathbb{1}_{x \geq 0}$.

6.3. **Ordinary differential equations on $p^{n,k}$, $v^{n,k}$ and $T^{n,k}$ in the exponential case.** From the functional equations (27), (41) and (43), we show that the functions $p^{n,k}$, $v^{n,k}$ and $T^{n,k}$ on $[0,a]$ are solutions of linear ordinary differential equations, in the exponential case.

Proposition 6.4. *Let n and $k \in \{1, \ldots, n-2\}$ be fixed and let us assume that $X \sim \mathcal{E}(1)$. There exist real numbers $\mu^{k,n}$ and $(r_m^{k,n})_{0 \leq m \leq k-1}$, depending only on n and k, such that $p^{n,k}$, $v^{n,k}$ and $T^{n,k}$ satisfy the*

following Linear Ordinary Differential Equations (ODEs) of order k: for $x \in [0,a]$:

$$\frac{d^k}{dx^k}p^{n,k}(x) = \left(1-\frac{k}{n}\right)\mu^{n,k}p^{n,k}(x) + \sum_{m=0}^{k-1} r_m^{n,k}\frac{d^m}{dx^m}p^{n,k}(x), \tag{45}$$

$$\frac{d^k}{dx^k}v^{n,k}(x) = \left(1-\frac{k}{n}\right)^2\mu^{n,k}v^{n,k}(x) + \sum_{m=0}^{k-1} r_m^{n,k}\frac{d^m}{dx^m}v^{n,k}(x), \tag{46}$$

$$\frac{d^k}{dx^k}T^{n,k}(x) = \sum_{m=1}^{k-1} r_m^{n,k}\frac{d^m}{dx^m}T^{n,k}(x) + \mu^{n,k}. \tag{47}$$

Notice that in (47) *the summation starts at $m=1$, while in* (45) *and* (46) *it starts at $m=0$. The coefficients $\mu^{k,n}$ and $(r_m^{k,n})_{0\le m\le k-1}$ are defined by a simple induction formula, see* (55).

Moreover, the functions $p^{n,k}$, $v^{n,k}$ and $T^{n,k}$ satisfy the following boundary conditions at point $x=a$: for $m \in \{0,\dots,k-1\}$

$$\frac{d^m}{dx^m}p^{n,k}(x)\Big|_{x=a} = 1, \tag{48}$$

$$\frac{d^m}{dx^m}v^{n,k}(x)\Big|_{x=a} = \frac{1}{n} + \left(1-\frac{1}{n}\right)2^m, \tag{49}$$

$$\frac{d^m}{dx^m}T^{n,k}(x)\Big|_{x=a} = \mathbb{1}_{m=0}. \tag{50}$$

The main tool for the proof of Proposition 6.4 is the following formula on the derivative of the density $f_{n,k}(.;x)$ with respect to x: for all $y > x$,

$$\begin{gathered}\frac{d}{dx}f_{n,1}(y;x) = nf_{n,1}(y;x)\\ \text{for } k\in\{2,\dots,n-1\},\ \frac{d}{dx}f_{n,k}(y;x) = (n-k+1)(f_{n,k}(y;x)-f_{n,k-1}(y;x)).\end{gathered} \tag{51}$$

Recall that $f(y;x) = \exp(-(y-x))$ for $y \ge x$. The proof of the first formula in (51) is straightforward, since $f_{n,1}(y;x) = n\exp(-n(y-x))$. For $k \in \{2,\dots,n-1\}$, we write (using (23))

$$\begin{aligned}\frac{d}{dx}f_{n,k}(y;x) &= \frac{d}{dx}\left(k\binom{n}{k}F(y;x)^{k-1}f(y;x)(1-F(y;x))^{n-k}\right)\\ &= k\binom{n}{k}\frac{d}{dx}\left((1-\exp(x-y))^{k-1}\exp\left((n-k+1)(x-y)\right)\right)\\ &= k\binom{n}{k}\Bigg(-(k-1)\exp(x-y)(1-\exp(x-y))^{k-2}\exp\left((n-k+1)(x-y)\right)\\ &\qquad +(n-k+1)(1-\exp(x-y))^{k-1}\exp\left((n-k+1)(x-y)\right)\Bigg)\\ &= (n-k+1)f_{n,k}(y;x) - (k-1)\frac{k\binom{n}{k}}{(k-1)\binom{n}{k-1}}f_{n,k-1}(y;x)\\ &= (n-k+1)\left(f_{n,k}(y;x)-f_{n,k-1}(y;x)\right).\end{aligned}$$

Remark 6.5. *A generalization of* (51) *holds in a more general case than the exponential setting. Indeed, if X has a density f, then, for $y > x$*

$$\begin{gathered}\frac{d}{dx}f_{n,1}(y;x) = \frac{f(x)}{1-F(x)}nf_{n,1}(y;x)\\ \textit{for } k\in\{2,\dots,n-1\},\ \frac{d}{dx}f_{n,k}(y;x) = \frac{f(x)}{1-F(x)}(n-k+1)\left(f_{n,k}(y;x)-f_{n,k-1}(y;x)\right).\end{gathered} \tag{52}$$

In the exponential case, the simplification $\frac{f(x)}{1-F(x)} = 1$ helps getting simpler formulae, which lead to the linear ODEs of Proposition 6.4. It is also worth noting the following formula $\frac{f(x)}{1-F(x)} = -\frac{d}{dx}\log(1-F(x)) = \frac{d}{dx}\Lambda(x)$, which explains the role played by the change of variable using the function Λ to reduce the general case to the exponential case.

Proof of Proposition 6.4: We mainly focus on the derivation of the ODE (45) for $p^{n,k}$. The ODEs (46) and (47) are obtained with similar arguments.

For any $1 \le l \le k$, we define for $0 \le x \le a$

$$I_l^{n,k}(x) = \int_x^a \left(1 - \frac{k}{n}\right) p^{n,k}(y) f_{n,l}(y;x)\, dy. \tag{53}$$

We also set $I_0^{n,k}(x) = (1 - \frac{k}{n})p^{n,k}(x)$.

As a consequence of (51), we get for $1 \le l \le k$

$$\frac{d}{dx} I_l^{n,k}(x) = (n - l + 1)(I_l^{n,k}(x) - I_{l-1}^{n,k}(x)).$$

Precisely, for $2 \le l \le k$, this formula directly follows from (53) and $f_{n,l}(x;x) = 0$. When $l = 1$,

$$\begin{aligned}\frac{d}{dx} I_1^{n,k}(x) &= -\left(1 - \frac{k}{n}\right) p^{n,k}(x) f_{n,1}(x;x) + n I_1^{n,k}(x) \\ &= n(I_1^{n,k}(x) - I_0^{n,k}(x)).\end{aligned}$$

The ODE (45) on $p^{n,k}$ is then obtained as follows.

- $p^{n,k}(x) - \theta_p^{n,k}(x) = I_k^{n,k}(x)$ (this is the functional equation (27)).
- For any $0 \le l \le k$, we prove by induction that the following formula holds:

$$\frac{d^l}{dx^l}\left(p^{n,k}(x) - \theta_p^{n,k}(x)\right) = \mu_l^{n,k} I_l^{n,k}(x) + \sum_{m=0}^{l-1} r_{m,l}^{n,k} \frac{d^m}{dx^m}\left(p^{n,k}(x) - \theta_p^{n,k}(x)\right). \tag{54}$$

The coefficients are defined recursively as follows:

$$\begin{aligned}&\mu_0^{n,k} = 1, \mu_{l+1}^{n,k} = -(n - k + l + 1)\mu_l^{n,k}; \\ &\begin{cases} r_{0,l+1}^{n,k} = -(n-k+l+1) r_{0,l}^{n,k}, & \text{if } l > 0, \\ r_{m,l+1}^{n,k} = r_{m-1,l}^{n,k} - (n-k+l+1) r_{m,l}^{n,k}, & 1 \le m \le l-1, \\ r_{l,l+1}^{n,k} = n - k + l + 1. \end{cases}\end{aligned} \tag{55}$$

- The ODE is obtained at $l = k$, using the definition of $I_0^{n,k}$ and setting $\mu^{n,k} = \mu_k^{n,k}$ and $r_m^{n,k} = r_{m,k}^{n,k}$:

$$\frac{d^k}{dx^k}\left(p^{n,k}(x) - \theta_p^{n,k}(x)\right) = \left(1 - \frac{k}{n}\right)\mu^{n,k} p^{n,k}(x) + \sum_{m=0}^{k-1} r_m^{n,k} \frac{d^m}{dx^m}\left(p^{n,k}(x) - \theta_p^{n,k}(x)\right).$$

Similarly, we obtain

$$\frac{d^k}{dx^k}\left(v^{n,k}(x) - \theta_v^{n,k}(x)\right) = \left(1 - \frac{k}{n}\right)^2 \mu^{n,k} v^{n,k}(x) + \sum_{m=0}^{k-1} r_m^{n,k} \frac{d^m}{dx^m}\left(v^{n,k}(x) - \theta_v^{n,k}(x)\right)$$

and

$$\frac{d^k}{dx^k} T^{n,k}(x) = \mu^{n,k} T^{n,k}(x) - r_0^{n,k} + \sum_{m=0}^{k-1} r_m^{n,k} \frac{d^m}{dx^m} T^{n,k}(x).$$

Since $r_0^{n,k} = (-1)^{k-1} n \dots (n-k+1) = -\mu^{n,k}$, after simplifications we obtain (47).

To obtain the ODEs given in Proposition 6.4, it remains to prove that

$$\begin{aligned}\frac{d^k}{dx^k}\theta_p^{n,k}(x) &= \sum_{m=0}^{k-1} r_m^{n,k}\frac{d^m}{dx^m}\theta_p^{n,k}(x),\\ \frac{d^k}{dx^k}\theta_v^{n,k}(x) &= \sum_{m=0}^{k-1} r_m^{n,k}\frac{d^m}{dx^m}\theta_v^{n,k}(x).\end{aligned} \tag{56}$$

The argument is as follows. Using Lemma 4.3, in the special case of exponential random variables, elementary computations show that for $0 \le x \le a$

$$\theta_p^{n,k}(x) = \sum_{j=0}^{k-1} k\binom{n-1}{k}\binom{k-1}{j}\frac{(-1)^j}{n-k+j}\exp\left((n-k+j+1)(x-a)\right).$$

This shows that $\theta_p^{n,k}$ as well as its derivatives, are linear combinations of the linearly independent functions $x \mapsto \exp((n-k+1)x)$, …, $x \mapsto \exp(nx)$. For our purpose, the exact expression of the coefficients does not matter.

But from Theorem 4.1, we know that $p^{n,k}(x) = P(x) = \exp(x-a)$. We thus conclude by a linear independence argument that

$$\begin{aligned}\frac{d^k}{dx^k}p^{n,k}(x) &= \left(1-\frac{k}{n}\right)\mu^{n,k}p^{n,k}(x) + \sum_{m=0}^{k-1} r_m^{n,k}\frac{d^m}{dx^m}p^{n,k}(x),\\ \frac{d^k}{dx^k}\theta_p^{n,k}(x) &= \sum_{m=0}^{k-1} r_m^{n,k}\frac{d^m}{dx^m}\theta_p^{n,k}(x).\end{aligned}$$

In particular, the second equality implies that for $0 \le j \le k-1$

$$\frac{d^k}{dx^k}\exp\left((n-k+j+1)(x-a)\right) = \sum_{m=0}^{k-1} r_m^{n,k}\frac{d^m}{dx^m}\exp\left((n-k+j+1)(x-a)\right). \tag{57}$$

The second equality of (56) can be proven using the same arguments.

Remark 6.6. *It is actually possible to prove directly the identity* (57) *from the definition of the coefficients* $r_m^{n,k}$*, without resorting to the result of Theorem 4.1. Indeed, let us introduce* $S_l^j := (n-k+j+1)^l - \sum_{m=0}^{l-1} r_{m,l}^{n,k}(n-k+j+1)^m$*. Thanks to the recursion formula* (55)*, introducing an appropriate telescoping sum, one obtains that the coefficients* (S_l^j) *satisfy* $S_0^j = 1$ *and* $S_{l+1}^j = (j-l)S_l^j$ *which implies that* $S_k^j = S_{j+1}^j = 0$ *for* $0 \le j \le k-1$*. One thus obtains that* $p^{n,k}$ *satisfies the ODE* (45) *without using Theorem 4.1.*

This approach actually yields an alternative proof of Theorem 4.1 in the exponential case, that can then be extended to the general case using the reduction to the exponential case explained in Section 3.2. For this, we check that $x \mapsto \exp(x-a)$ *is solution of the differential equation on* $p^{n,k}$*. First, it satisfies the appropriate condition at* $x = a$ *given in Proposition 6.4. Second, we observe that* $t = 1$ *is a root of the characteristic polynomial equation associated with the linear ODE which writes (the unknown variable being* t*):*

$$\frac{(n-t)\dots(n-k+1-t)}{n\dots(n-k+1)} = \frac{(n-k)}{n}.$$

See (59) *below for the derivation of the caracteristic equation.*

To conclude the proof of Proposition 6.4, it remains to show the boundary conditions at $x = a$. From the recursion equation (54) leading to the differential equation on $p^{n,k}$, and Lemma 4.3, we have for $0 \leq l \leq k-1$

$$\begin{aligned}
\frac{d^l}{dx^l} p^{n,k}(x)\Big|_{x=a} &= \frac{d^l}{dx^l} \theta_p^{n,k}(x)\Big|_{x=a} \\
&= \frac{d^l}{dx^l} \left(\exp(x-a)(1 - F_{n-1,k}(a;x))\right)\Big|_{x=a} \\
&= 1 + \sum_{j=1}^{l} \binom{l}{j} \frac{d^j}{dx^j}(1 - F_{n-1,k}(a;x))\Big|_{x=a} \\
&= 1 - \sum_{j=1}^{l} \binom{l}{j} \frac{d^{j-1}}{dx^{j-1}} f_{n-1,k}(a;x)\Big|_{x=a} \\
&= 1.
\end{aligned}$$

Indeed, the second identity of (51) yields $\frac{d^{j-1}}{dx^{j-1}} f_{n-1,k}(a;x)\Big|_{x=a} = 0$ for any $1 \leq j \leq k-1$.

Similarly, from the derivation of the ODE on $v^{n,k}$ we have

$$\begin{aligned}
\frac{d^l}{dx^l} v^{n,k}(x)\Big|_{x=a} &= \frac{d^l}{dx^l} \theta_v^{n,k}(x)\Big|_{x=a} \\
&= \frac{1}{n} \frac{d^l}{dx^l} \left(\exp(x-a)\left(1 - F_{n-1,k}(a;x)\right)\right)\Big|_{x=a} \\
&\quad + \left(1 - \frac{1}{n}\right) \frac{d^l}{dx^l} \left(\exp(2(x-a))\left(1 - F_{n-2,k}(a;x)\right)\right)\Big|_{x=a} \\
&= \frac{1}{n} + \left(1 - \frac{1}{n}\right) 2^l,
\end{aligned}$$

using Lemma 6.3, and the same arguments as for $p^{n,k}$.

Finally, using similar arguments,

$$\begin{aligned}
T^{n,k}(a) &= 1; \\
\frac{d^l}{dx^l} T^{n,k}(x)\Big|_{x=a} &= \frac{d^l}{dx^l} 1\Big|_{x=a} = 0, \quad 1 \leq l \leq k-1.
\end{aligned}$$

This concludes the proof of Proposition 6.4. □

In the next Sections, we analyze the differential equations on $v^{n,k}$ and $T^{n,k}$. We are not able to derive explicit expressions for the solutions, except when $k = 1$ (see Remark 5.6). However, we are able to analyze quantitatively the behavior when $n \to +\infty$.

6.4. **Asymptotic expansion for the variance.** In this Section, we prove Proposition 5.2 in the exponential case, namely (39). Since

$$\mathrm{Var}(\hat{p}^{n,k}(x)) = v^{n,k}(x) - (p^{n,k}(x))^2,$$

and we know from Theorem 4.1 that $p^{n,k}(x) = P(x) = \exp(x-a)$, we focus on $v^{n,k}$. This function is solution of the linear ODE of order k given in Proposition 6.4, which we rewrite here:

$$\frac{d^k}{dx^k} v^{n,k}(x) = \left(1 - \frac{k}{n}\right)^2 \mu^{n,k} v^{n,k}(x) + \sum_{m=0}^{k-1} r_m^{n,k} \frac{d^m}{dx^m} v^{n,k}(x).$$

To understand how the solution $v^{n,k}$ behaves, we thus need to study the following associated polynomial equation with unknown t:

$$t^k - \sum_{m=0}^{k-1} r_m^{n,k}\, t^m - \mu^{n,k} \left(1 - \frac{k}{n}\right)^2 = 0.$$

In order to study the behavior of the (complex) roots of this equation, we first observe that

$$t^k - \sum_{m=0}^{k-1} r_m^{n,k}\, t^m = (t-n)\dots(t-n+k-1), \tag{58}$$

since by (57) the k roots of this polynomial function are n, …, $n-k+1$.

Moreover, $\mu^{n,k} = (-1)^k n \dots (n-k+1)$; therefore, the polynomial equation can be rewritten

$$\frac{(n-t)\dots(n-k+1-t)}{n\dots(n-k+1)} = \frac{(n-k)^2}{n^2}. \tag{59}$$

Proposition 6.7. *Let k be fixed. There is a unique root of (59) in the real interval $[1,2]$, denoted by $\beta^1_{n,k}$. The other roots of (59) in $\mathbb{C}$ are denoted by $\beta^2_{n,k}, \dots, \beta^k_{n,k}$.Moreover, we have the following asymptotic expansions when $n \to +\infty$:*

$$\beta^1_{n,k} = 2 - \frac{1}{n} - \frac{k-1}{2n^2} + \mathrm{o}\left(\frac{1}{n^2}\right)$$
$$\beta^l_{n,k} \sim n\left(1 - \exp(i2\pi(l-1)/k)\right), \text{ for } l \in \{2,\dots,k\}.$$

Proof of Proposition 6.7: We observe that the complex numbers $\left(\frac{\beta^l_{n,k}}{n}\right)_{1\le l\le k}$ are solutions of the following polynomial equation of degree k (with unknown $\tilde{t}$):

$$\frac{(1-\tilde{t})\dots(1-\frac{k-1}{n}-\tilde{t})}{1\dots(1-\frac{k-1}{n})} = \left(1-\frac{k}{n}\right)^2.$$

The claim for $l \in \{2,\dots,k\}$ then follows by continuity of the roots of a polynomial function of fixed degree with respect to the coefficients. Indeed, in the limit $n \to +\infty$ we obtain the equation $(1-\tilde{t})^k = 1$, whose roots are $(1-\exp(i2\pi(l-1)/k))_{1\le l\le k}$. If $2 \le l \le k$, we thus obtain $\beta^l_{n,k} \sim n\left(1-\exp(i2\pi(l-1)/k)\right)$. This argument for $l=1$ yields that $\beta^1_{n,k}/n$ goes to 0, and this is not sufficient for our purposes.

To study the behavior of $\beta^1_{n,k}$, let us introduce the polynomial function

$$P_{n,k}(t) = \frac{(n-t)\dots(n-k+1-t)}{n\dots(n-k+1)}.$$

This function is strictly non-decreasing on the interval $(-\infty, n-k+1]$ (which contains no root of $P_{n,k}$ and of its derivative).

Now straightforward computations show that $P_{n,k}(1) > (1-\frac{k}{n})^2 > P_{n,k}(2)$, and thus $P_{n,k}$ admits a root in the interval $[1,2]$. For n sufficiently large, since $\beta^l_{n,k} \notin [1,2]$, for $l \ge 1$, we get $\beta^1_{n,k} \in [1,2]$.

Moreover, elementary computations show that

$$P_{n+1,k}(\beta^1_{n,k}) = \frac{(n+1-\beta^1_{n,k})(n-k)}{(n-k+1-\beta^1_{n,k})n}\big(1-\frac{k}{n+1}\big) > \big(1-\frac{k}{n+1}\big),$$

and therefore since $P_{n+1,k}$ is non-decreasing, $\beta^1_{n,k} < \beta^1_{n+1,k}$: the sequence $(\beta_{n,k})_{n>k}$ is non-decreasing. Since $\beta_{n,k} \le 2$ for any $n > k$, the sequence converges.

To identify the limit, and higher order terms in the expansion of $\beta^1_{n,k}$, one postulates an ansatz $\beta^1_{n,k} = \beta_{\infty,k} - \frac{\beta_{\infty,1,k}}{n} - \frac{\beta_{\infty,2,k}}{n} + \mathrm{o}\left(\frac{1}{n^2}\right)$, and identifies successively $\beta_{\infty,k} = 2$, $\beta_{\infty,1,k} = 1$, $\beta_{\infty,2,k} = \frac{k-1}{2}$. □

For n large enough, all the roots are therefore simple, and we can express the function $v^{n,k}$ in the following way: for $0 \le x \le a$

$$v^{n,k}(x) = \sum_{l=1}^{k} \eta^l_{n,k} \exp\left(\beta^l_{n,k}(x-a)\right), \tag{60}$$

for some complex numbers $(\eta^l_{n,k})_{1\le l\le k}$, satisfying appropriate conditions to satisfy the boundary conditions (49). In particular, these complex numbers are such that $v^{n,k}$ is real-valued which implies necessarily $\eta^1_{n,k} \in \mathbb{R}$. Actually, these complex numbers are solution to a system of linear equations (which corresponds

to (49)). Using Cramer's rule, we give explicit formulae and then get asymptotic expansions for each $\eta^l_{n,k}$ when $n \to +\infty$.

Proposition 6.8. *Let k be fixed. When $n \to +\infty$, we have*

$$\eta^1_{n,k} = 1 + \mathrm{o}\left(\frac{1}{n}\right)$$
$$\eta^l_{n,k} = \mathrm{o}(1), \text{ for } k \in \{2, \dots, k\}.$$

Thanks to the previous result and to the expression of $v^{n,k}$ as a combination of exponential functions (60), subtracting $P(x)^2 = \exp(2(x-a))$, we obtain the desired asymptotic expansion (39) for the variance when $n \to +\infty$. Notice that since x is assumed to be strictly smaller than a, the only contribution which remains is $\eta^1_{n,k} \exp\left(\beta^1_{n,k}(x-a)\right)$. The other terms in the sum (60) vanish exponentially fast. This concludes the proof of Proposition 5.2.

We end this Section with the proof of Proposition 6.8.

Proof of Proposition 6.8: The family $(\eta^l_{n,k})_{1\le l\le k}$ is solution of the following system of linear equations (using (49)):

$$\begin{cases} \eta^1_{n,k} + \eta^2_{n,k} + \ldots + \eta^k_{n,k} = v^{n,k}(a) = 1 \\ \eta^1_{n,k}\beta^1_{n,k} + \eta^2_{n,k}\beta^2_{n,k} + \ldots + \eta^k_{n,k}\beta^k_{n,k} = \frac{d}{dx}v^{n,k}(x)\Big|_{x=a} = 2 - \frac{1}{n} \\ \ldots \\ \eta^1_{n,k}(\beta^1_{n,k})^{k-1} + \eta^2_{n,k}(\beta^2_{n,k})^{k-1} + \ldots + \eta^k_{n,k}(\beta^k_{n,k})^{k-1} = \frac{d^{k-1}}{dx^{k-1}}v^{n,k}(x)\Big|_{x=a} = \frac{1}{n} + \left(1 - \frac{1}{n}\right)2^{k-1}. \end{cases}$$

Using Cramer's rule (which gives the solution of an invertible linear system thanks to ratios of determinants), we see that

$$\eta^1_{n,k} = \frac{1}{n}\frac{V(1, \dots, \beta^k_{n,k})}{V(\beta^1_{n,k}, \dots, \beta^k_{n,k})} + \left(1 - \frac{1}{n}\right)\frac{V(2, \dots, \beta^k_{n,k})}{V(\beta^1_{n,k}, \dots, \beta^k_{n,k})},$$

where $V(\lambda_1, \dots, \lambda_k) = \det(\lambda_j^{i-1})_{1\le i,j\le k}$ denotes the Vandermonde determinant of the complex numbers $(\lambda_i)_{1\le i\le k}$. We recall that $V(\lambda_1, \dots, \lambda_k) = \prod_{1\le i<j\le k}(\lambda_j - \lambda_i)$.

Straightforward simplifications then imply that

$$\eta^1_{n,k} = \frac{1}{n}\frac{\prod_{2\le l\le k}(\beta^l_{n,k} - 1)}{\prod_{2\le l\le k}(\beta^l_{n,k} - \beta^1_{n,k})} + \left(1 - \frac{1}{n}\right)\frac{\prod_{2\le l\le k}(\beta^l_{n,k} - 2)}{\prod_{2\le l\le k}(\beta^l_{n,k} - \beta^1_{n,k})}.$$

The asymptotic results on the coefficients $\beta^l_{n,k}$ given in Proposition 6.7 finally show that $\eta^1_{n,k} = 1 + \mathrm{o}\left(\frac{1}{n}\right)$.

The proof that $\eta^l_{n,k} = o(1)$ for $k \in \{2, \dots, k\}$ follows the same lines. □

6.5. **Asymptotic expansion for $T^{n;k}$.** In this Section, we prove Proposition 5.3 in the exponential case, namely (40), following the same approach as in the previous Section. Recall from Proposition 6.4 that $T^{n,k}$ is solution of the following linear differential equation of order k:

$$\frac{d^k}{dx^k}T^{n,k}(x) - \sum_{m=1}^{k-1} r^{n,k}_m \frac{d^m}{dx^m}T^{n,k}(x) = \mu^{n,k}.$$

The associated polynomial equation $y^k - \sum_{m=1}^{k-1} r^{n,k}_m\, y^m = 0$ admits 0 as a root. Moreover, thanks to (58) it can be rewritten as

$$\frac{(n-t)\dots(n-k+1-t)}{n\dots(n-k+1)} = 1. \tag{61}$$

This formulation allows to get the following analog of Proposition 6.7, on the k roots $(\alpha^l_{n,k})_{1\le l\le k} \in \mathbb{C}^k$ of (61).

Proposition 6.9. *Let k be fixed. When $n \to +\infty$, the roots $(\alpha^l_{n,k})_{1\le l\le k}$ of (61) satisfy:*

$$\alpha^1_{n,k} = 0$$
$$\alpha^l_{n,k} \sim n\left(1 - \exp(i2\pi(l-1)/k)\right), \text{ for } l \in \{2, \dots k\}.$$

We omit the proof of Proposition 6.9, since it is very similar to the proof of Proposition 6.7. We have already identified that $\alpha^1_{n,k} = 0$ and the asymptotic formulae for $\alpha^l_{n,k}$ when $l \in \{2, \dots, k\}$ are obtained with exactly the same arguments as for $\beta^l_{n,k}$ in Proposition 6.7. Again, for n large enough, the roots $(\alpha^l_{n,k})_{l\in\{1,\dots,k\}}$ are pairwise distinct.

The two differences with the analysis performed in the previous Section are the following. First, the differential equation (47) on $T^{n,k}$ contains a non-zero right-hand side (namely a constant). Moreover, constant functions are solutions of the differential equation without this right-hand side, since 0 is a root of (61).

Therefore $T^{n,k}$ can be expressed as a sum of an affine function and of exponential functions, for n large enough (so that roots are pairwise distinct): for any $0 \le x \le a$

$$T^{n,k}(x) = \Delta_{n,k}(a-x) + \delta^1_{n,k} + \sum_{l=2}^{k} \delta^l_{n,k} \exp(\alpha^l_{n,k}(x-a)), \tag{62}$$

for some complex coefficients $\Delta_{n,k}$ and $\delta^l_{n,k}$, for $1 \le l \le k$.

We prove the following result, which then yields Proposition 5.3:

Proposition 6.10. *Let k be fixed. When $n \to +\infty$, we have*

$$\Delta_{n,k} = n\left(\frac{1}{k} - \frac{k-1}{2kn} + \mathrm{o}\left(\frac{1}{n}\right)\right),$$
$$\delta^1_{n,k} = \frac{3k-1}{2k} + \mathrm{o}(1) = n\left(\frac{3k-1}{2kn} + \mathrm{o}\left(\frac{1}{n}\right)\right),$$
$$\delta^l_{n,k} = O(1), \quad l \in \{2, \dots k\}.$$

Proof: First, the differential equation (47) on $T^{n,k}$ in Proposition 6.4 is in fact valid on $(-\infty, a]$, not only on $[0, a]$. It corresponds to the estimation by the AMS algorithm of $P(x) = \mathbb{P}(x + X > a) = \exp(x - a)$ for $x \le a$ and $X \sim \mathcal{E}(1)$ exponentially distributed.

Inserting (62) into the functional equation (43) and letting $x \to -\infty$ yields

$$\Delta_{n,k} = \frac{1}{\int_0^{+\infty} z\, f_{n,k}(z;0)dz}.$$

Indeed, we get, using the change of variable $z = y - x$ and the fact that $f_{n,k}(y;x) = f_{n,k}(y-x;0)$

$$\begin{aligned}
T^{n,k}(x) &= 1 + \int_x^a \Delta_{n,k}(x-y) f_{n,k}(y;x)dy + \int_x^a \Delta_{n,k}(a-x) f_{n,k}(y;x)dy \\
&\quad + \delta^1_{n,k} \int_x^a f_{n,k}(y;x)dy + \sum_{l=2}^{k} \delta^l_{n,k} \int_x^a e^{\alpha^l_{n,k}(y-a)} f_{n,k}(y;x)dy \\
&= 1 - \Delta_{n,k} \int_0^{a-x} z\, f_{n,k}(z;0)dz + \Delta_{n,k}(a-x) \int_0^{a-x} f_{n,k}(z;0)dz \\
&\quad + \delta^1_{n,k} \int_0^{a-x} f_{n,k}(z;0)dz + \sum_{l=2}^{k} \delta^l_{n,k} e^{\alpha^l_{n,k}(x-a)} \int_0^{a-x} e^{\alpha^l_{n,k} z} f_{n,k}(z;0)dz.
\end{aligned}$$

Taking the limit $x \to -\infty$ above and in (62), we obtain

$$T^{n,k}(x) = \Delta_{n,k}(a-x) + \delta^1_{n,k} + \mathrm{o}(1),$$
$$T^{n,k}(x) = \Delta_{n,k}(a-x) + \delta^1_{n,k} + 1 - \Delta_{n,k} \int_0^{a-x} z\, f_{n,k}(z;0)dz + \mathrm{o}(1),$$

and thus $1 - \Delta_{n,k} \int_0^{+\infty} f_{n,k}(z;0)dz = 0$.

Let us define $M_{n,k} = \int_0^{+\infty} z f_{n,k}(z;0)dz$. Then, using (51) and an integration by parts, we get

$$M_{n,1} = \frac{1}{n},$$
$$M_{n,k} = M_{n,k-1} + \frac{1}{n-k+1}.$$

Indeed,

$$\begin{aligned} M_{n,k} - M_{n,k-1} &= \int_0^{+\infty} z f_{n,k}(z;0) dz \\ &= \int_0^{+\infty} z \left(f_{n,k}(z;0) - f_{n,k}(z;0) \right) dz \\ &= -\frac{1}{n-k+1} \int_0^{+\infty} z \frac{d}{dz} f_{n,k}(z;0) dz \\ &= \frac{1}{n-k+1} \int_0^{+\infty} f_{n,k}(z;0) dz = \frac{1}{n-k+1}. \end{aligned}$$

We therefore obtain

$$\begin{aligned} M_{n,k} &= \frac{1}{n} + \ldots + \frac{1}{n-k+1} \\ &= \frac{1}{n} \left(1 + \frac{1}{1-1/n} + \ldots + \frac{1}{1-(k-1)/n} \right) \\ &= \frac{1}{n} \left(k + \frac{k(k-1)}{2n} + \mathrm{o}(1/n) \right) \end{aligned}$$

and

$$\begin{aligned} \Delta_{n,k} &= \frac{1}{M_{n,k}} = \frac{n}{k} \frac{1}{1 + \frac{k-1}{2n} + \mathrm{o}(1/n)} \\ &= \frac{n}{k} \left(1 - \frac{k-1}{2n} + o(1/n) \right). \end{aligned}$$

The values of $\delta_{n,k}^l$ for $1 \leq l \leq k$ depend only on the conditions at $x = a$ for $T^{n,k}$ and its derivatives up to order $k-1$, namely (50). They are solutions of a system of linear equations and they can be expressed thanks to Cramer's rule. The family $(\delta_{n,k}^l)_{1 \leq l \leq k}$ is solution of the following system of linear equations:

$$\begin{cases} \delta_{n,k}^1 + \delta_{n,k}^2 + \ldots + \delta_{n,k}^k = T^{n,k}(a) = 1 \\ 0 + \delta_{n,k}^2 \alpha_{n,k}^2 + \ldots + \delta_{n,k}^k \alpha_{n,k}^k = \frac{d}{dx} T^{n,k}(x) \Big|_{x=a} + \Delta_{n,k} = \Delta_{n,k} \\ \ldots \\ 0 + \delta_{n,k}^2 (\alpha_{n,k}^2)^{k-1} + \ldots + \delta_{n,k}^k (\alpha_{n,k}^k)^{k-1} = \frac{d^{k-1}}{dx^{k-1}} T^{n,k}(x) \Big|_{x=a} = 0. \end{cases}$$

Using Cramer's rule and Vandermonde determinants, we see that

$$\delta_{n,k}^1 V(0, \alpha_{n,k}^2, \ldots, \alpha_{n,k}^k) = \det \begin{pmatrix} 1 & 1 & \ldots & 1 \\ \Delta_{n,k} & \alpha_{n,k}^2 & \ldots & \alpha_{n,k}^k \\ 0 & (\alpha_{n,k}^2)^2 & \ldots & (\alpha_{n,k}^k)^2 \\ \vdots & \vdots & \ldots & \vdots \\ 0 & (\alpha_{n,k}^2)^{k-1} & \ldots & (\alpha_{n,k}^k)^{k-1} \end{pmatrix}. \tag{63}$$

Let us introduce a few more notations. We define $\xi_{l,k} = \exp(i 2\pi (l-1)/k)$ for $2 \leq l \leq k$, and the polynomial function:

$$Q(z) = \det \begin{pmatrix} 1 & 1 & \ldots & 1 \\ -z & (1-\xi_{2,k}) & \ldots & (1-\xi_{k,k}) \\ z^2 & (1-\xi_{2,k})^2 & \ldots & (1-\xi_{k,k})^2 \\ \vdots & \vdots & \ldots & \vdots \\ (-z)^{k-1} & (1-\xi_{2,k})^{k-1} & \ldots & (1-\xi_{k,k})^{k-1} \end{pmatrix}.$$

Then by considering the limit $n \to +\infty$ in (63), plugging the asymptotic expansions for $\alpha_{n,k}^l$ and for $\Delta_{n,k}$ from Propositions 6.9 and 6.10, one obtains

$$\delta^1_{\infty,k} := \lim_{n\to+\infty} \delta^1_{n,k} = 1 + \frac{Q'(0)}{kQ(0)}.$$

Since $Q(z)$ is a Vandermonde determinant, we have the explicit formula

$$\begin{aligned} Q(z) &= V(-z, 1-\xi_{2,k}, \dots, 1-\xi_{k,k}) = V(1-\xi_{2,k}, \dots, 1-\xi_{k,k}) \prod_{l=2}^{k} (1-\xi_{l,k}+z) \\ &= V(1-\xi_{2,k}, \dots, 1-\xi_{k,k}) R(z+1), \end{aligned}$$

with $R(z) = \prod_{l=2}^{k}(z-\xi_{l,k}) = \frac{z^k-1}{z-1}$; thus $\delta^1_{\infty,k} = 1 + \frac{1}{k}\frac{R'(1)}{R(1)}$.

If we now write that $z^k - 1 = R(z)(z-1)$, differentiating and setting $z = 1$ we obtain $R(1) = k$; a similar argument shows that $R'(1) = \frac{k(k-1)}{2}$, so that finally

$$\delta^1_{\infty,k} = 1 + \frac{k-1}{2k} = \frac{3k-1}{2k}.$$

A similar Cramer's rule holds for each $\delta^l_{n,k}$ when $l \geq 2$. It is then easy to check that $\delta^l_{n,k} = O(1)$. In fact, an analytical formula for the limit of $\delta^l_{n,k}$ in the limit $n \to +\infty$ can be written, but we do not need such a sharp result for our purposes.

This concludes the proof of Proposition 6.9. □

References

[1] S. Asmussen, P. W. Glynn. Stochastic simulation: Algorithms and analysis Springer, 2007.

[2] C.E. Bréhier, M. Gazeau, L. Goudenège, T. Lelièvre, and M. Rousset. Mathematical analysis of adaptive multilevel splitting algorithms. In preparation.

[3] F. Cérou, P. Del Moral, T. Furon, and A. Guyader. Sequential Monte Carlo for rare event estimation. *Stat. Comput.*, 22(3):795–808, 2012.

[4] F. Cérou and A. Guyader. Adaptive multilevel splitting for rare event analysis. *Stoch. Anal. Appl.*, 25(2):417–443, 2007.

[5] F. Cérou and A. Guyader. Adaptive particle techniques and rare event estimation. In *Conference Oxford sur les méthodes de Monte Carlo séquentielles*, volume 19 of *ESAIM Proc.*, pages 65–72. EDP Sci., Les Ulis, 2007.

[6] F. Cérou, A. Guyader, T. Lelièvre, and D. Pommier. A multiple replica approach to simulate reactive trajectories. *J. Chem. Phys.*, 134:054108, 2011.

[7] A. Guyader, N. Hengartner, and E. Matzner-Løber. Simulation and estimation of extreme quantiles and extreme probabilities. *Appl. Math. Optim.*, 64(2):171–196, 2011.

[8] J.M. Hammersley and D.C. Handscomb. *Monte Carlo methods*. Methuen & Co. Ltd., 1965.

Université Paris-Est, CERMICS (ENPC), INRIA, 6-8-10 avenue Blaise Pascal, 77455 Marne-la-Vallée, France

INRIA Paris-Rocquencourt, Domaine de Voluceau - Rocquencourt, B.P. 105 - 78153 Le Chesnay, France

E-mail address: brehierc@cermics.enpc.fr

E-mail address: lelievre@cermics.enpc.fr

E-mail address: mathias.rousset@inria.fr