Central Limit Theorem for Adaptive Multilevel Splitting Estimators in an Idealized Setting

Charles-Edouard Bréhier, Ludovic Goudenège, and Loïc Tudela

Abstract The Adaptive Multilevel Splitting (AMS) algorithm is a powerful and versatile iterative method to estimate the probabilities of rare events.
We prove a new central limit theorem for the associated AMS estimators introduced in [5], and which have been recently revisited in [3] – the main result there being (non-asymptotic) unbiasedness of the estimators. To prove asymptotic normality, we rely on and extend the technique presented in [3]: the (asymptotic) analysis of an integral equation.
Numerical simulations illustrate the convergence and the construction of Gaussian confidence intervals.

Key words: Monte-Carlo simulation, rare events, multilevel splitting, central limit theorem

Subject Classifications: 65C05; 65C35; 60F05

1 Introduction

Many models from physics, chemistry or biology involve stochastic systems for different purposes: taking into account uncertainty with respect to data parameters,

---

Charles-Edouard Bréhier
Université Paris-Est, CERMICS (ENPC), 6-8-10 Avenue Blaise Pascal, Cité Descartes , F-77455 Marne-la-Vallée, France
INRIA Paris-Rocquencourt, Domaine de Voluceau - Rocquencourt, B.P. 105 - 78153 Le Chesnay, France, e-mail: brehierc@cermics.enpc.fr

Ludovic Goudenège
Fédération de Mathématiques de l'École Centrale Paris, CNRS, Grande voie des vignes, 92295 Châtenay-Malabry, France e-mail: goudenege@math.cnrs.fr

Loïc Tudela
Ensae ParisTech, 3 Avenue Pierre Larousse, 92240 Malakoff, France e-mail: loic.tudela@ensae-paristech.fr

or allowing for dynamical phase transitions between different configurations of the system. This phenomenon often referred to as metastability is observed, for instance, when one studies a d-dimensional overdamped Langevin dynamics:

$$dX_t = -\nabla V(X_t)dt + \sqrt{2\beta^{-1}}dW_t,$$

associated with a potential function V with several local minima. Here W denotes a d-dimensional standard Wiener process. When the inverse temperature β increases, the transitions become rare events (their probability decreases exponentially fast).

In this paper, we adopt a numerical point of view, and analyze a method which outperforms a pure Monte-Carlo method for a given computational effort in the small probability regime (in terms of relative error). Two important families of methods have been introduced in the 1950s and next have been extensively developed, in order to efficiently address this rare event estimation problem: importance sampling, and importance/multilevel splitting - see [11], and [9] for a more recent treatment. We refer for instance to [12] for a more general presentation.

The method we study in this work is a multilevel splitting algorithm. The main advantage of this kind of methods is that they are non-intrusive: the model does not need to be modified in order to obtain a more efficient Monte-Carlo method. The method we study has an additional feature: adaptive computations (of levels) are made on-the-fly. To explain more precisely the algorithm and its properties, from now on we only focus on a simpler, generic setting for the rare event estimation problem.

Let X be a real random variable, and a be a given threshold. We want to estimate the tail probability $p := \mathbb{P}(X > a)$. The splitting strategy, in the regime when a becomes large, consists in introducing the following decomposition of p, as a product of conditional probabilities:

$$\mathbb{P}(X > a) = \mathbb{P}(X > a_n | X > a_{n-1}) \dots \mathbb{P}(X > a_2 | X > a_1)\mathbb{P}(X > a_1),$$

for a sequence of levels $a_1 < \dots < a_{n-1} < a_n = a$. The common interpretation of this formula is that the event that $X > a$ is split in n conditional probabilities for X, which are each much larger than p, and are thus easier to estimate.

To optimize the variance, the levels must be chosen such that all the conditional probabilities are equal to $p^{1/n}$, with n as large as possible. However, levels satisfying this condition are not known *a priori* in practical cases.

Notice that, in principle, to apply this splitting strategy, one needs to know how to sample according to the conditional distributions appearing in the splitting formula. If this condition holds, we say that we are in an idealized setting.

Adaptive techniques based on multilevel splitting, where the levels are computed on-the-fly, have been introduced in the 2000s in various contexts, under different names: Adaptive Multilevel Splitting (AMS) [5], [6], [7], Subset simulation [2] and Nested sampling [13] for instance.

In this paper, we focus on the versions of AMS algorithms studied in [3], following [5]. Such algorithms depend on two parameters: a number of (interacting) replicas

n, and a fixed integer $k \in \{1, \ldots, n-1\}$, such that a proportion k/n of replicas are killed and resampled at each iteration. The version with $k = 1$ has been studied in [10], and is also (in the idealized setting) a special case of the Adaptive Last Particle Algorithm of [14].

A family of estimators $(\hat{p}^{n,k})_{n \geq 2, 1 \leq k \leq n-1}$ is introduced in [3] – see (2) and (3). The main property established there is *unbiasedness*: for all values n and k the equality $\mathbb{E}[\hat{p}^{n,k}] = p$ holds true – note that this statement is not an asymptotic result. Moreover, an analysis of the computational cost is provided there, in the regime $n \to +\infty$, with fixed k. However, comparisons, when k changes, are made using a cumbersome procedure: M independent realizations of the algorithm are necessary to define a new estimator, as an empirical mean of $\hat{p}_1^{n,k}, \ldots, \hat{p}_M^{n,k}$, and finally one studies the limit when $M \to +\infty$. The aim of this paper is to remove this procedure: we prove directly an asymptotic normality result for the estimator $\hat{p}^{n,k}$, when $n \to +\infty$, with fixed k. Such a result allows to directly rely on asymptotic Gaussian confidence intervals.

Note that other Central Limit Theorems for Adaptive Multilevel Splitting estimators (in different parameter regimes for n and k) have been obtained in [4, 5, 8].

The main result of this paper is Theorem 1: if k and a are fixed, under the assumption that the cumulative distribution function of X is continuous, when $n \to +\infty$, the random variable $\sqrt{n}\big(\hat{p}^{n,k} - p\big)$ converges in law to a centered Gaussian random variable, with variance $-p^2 \log(p)$ (independent of k).

The main novelty of the paper is the treatment of the case $k > 1$: indeed when $k = 1$ (see [10]) the law of the estimator is explicitly known (it involves a Poisson random variable with parameter $-n \log(p)$): the asymptotic normality of $\log(\hat{p}^{n,1})$ is a consequence of straightforward computation, and the central limit theorem for $\hat{p}^{n,1}$ easily follows using the delta-method. When $k > 1$, the law is more complicated and not explicitly known; the key idea is to prove that the characteristic function of $\log(\hat{p}^{n,k})$ satisfies a functional equation, following the strategy in [3]; the basic ingredient is a decomposition according to the first step of the algorithm.

One of the main messages of this paper is thus that the functional equation technique is a powerful tool in order to prove several key properties of the AMS algorithm in the idealized setting: unbiasedness and asymptotic normality.

The paper is organized as follows. In Section 2, we introduce the main objects: the idealized setting (Section 2.1) and the AMS algorithm (Section 2.2). Our main result (Theorem 1) is stated in Section 2.3. Section 3 is devoted to the detailed proof of this result. Finally Section 4 contains a numerical illustration of the Theorem.

2 Adaptive Multilevel Splitting Algorithms

2.1 Setting

Let X be a real random variable. We assume that $X > 0$ almost surely. The aim is the estimation of the probability $p = \mathbb{P}(X > a)$, where $a > 0$ is a threshold. When a

goes to $+\infty$, p goes to 0. More generally, we introduce the conditional probability for $0 \leq x \leq a$

$$P(x) = \mathbb{P}(X > a | X > x). \tag{1}$$

Note that the quantity of interest satisfies $p = P(0)$; moreover $P(a) = 1$.

Let F denote the cumulative distribution function of X: $F(x) = \mathbb{P}(X \leq x)$ $\forall x \in \mathbb{R}$.

The following standard assumption ([3, 5]) is crucial for the study in this paper.

Assumption 1 *The function F is assumed to be continuous.*

2.2 The AMS Algorithm

The algorithm depends on two parameters:

- the number of replicas $n \geq 2$;
- the number $k \in \{1, \ldots, n-1\}$ of replicas that are resampled at each iteration.

The other necessary parameters are the stopping threshold a and the initial condition $x \in [0, a]$. On the one hand, in practice, one applies the algorithm with $x = 0$ to estimate p. On the other hand, introducing an additional variable x for the initial condition is a key tool for the theoretical analysis of the algorithm.

In the sequel, when a random variable X_i^j is written, the subscript i denotes the index in $\{1, \ldots, n\}$ 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 $Y = (Y_1, \ldots, Y_n)$ independent and identically distributed (i.i.d.) real valued random variables with continuous cumulative distribution function, there exists almost surely a unique (random) permutation σ of $\{1, \ldots, n\}$ such that $Y_{\sigma(1)} < \ldots < Y_{\sigma(n)}$. For any $k \in \{1, \ldots, n\}$, we then use the classical notation $Y_{(k)} = Y_{\sigma(k)}$ to denote the k-th order statistics of the sample Y.

We are now in position to describe the Adaptive Multilevel Splitting (AMS) algorithm.

Algorithm 1 (Adaptive Multilevel Splitting)
Initialization: *Define $Z^0 = x$. Sample n i.i.d. realizations $X_1^0, \ldots, X_n^0$, with the law $\mathscr{L}(X | X > x)$.*

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

Set $j = 1$.

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

- *Conditionally on Z^j, sample k new independent random variables $(Y_1^j, \ldots, Y_k^j)$, according to the conditional distribution $\mathscr{L}(X | X > Z^j)$.*
- *Set*

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

In other words, the particle with index i is killed and resampled according to the law $\mathscr{L}(X|X > Z^j)$ 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 $i \in \left\{\sigma^j(1), \ldots, \sigma^j(k)\right\}$.

- *Define $Z^{j+1} = X^j_{(k)}$, the k-th order statistics of the sample $X^j = (X^j_1, \ldots, X^j_n)$, and σ^{j+1} the (a.s.) unique associated permutation: $X^j_{\sigma^{j+1}(1)} < \ldots < X^j_{\sigma^{j+1}(n)}$.*
- *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 $Z^{J^{n,k}(x)+1} \geq a$.*

For a schematic representation of the algorithm, we refer for instance to [5].

We are now in position to define the estimator $\hat{p}^{n,k}(x)$ of the probability $P(x)$:

$$\hat{p}^{n,k}(x) = C^{n,k}(x)\left(1 - \frac{k}{n}\right)^{J^{n,k}(x)}, \tag{2}$$

with

$$C^{n,k}(x) = \frac{1}{n}\mathrm{Card}\left\{i\,;X_i^{J^{n,k}(x)} \geq a\right\}. \tag{3}$$

When $x = 0$, to simplify notations we set $\hat{p}^{n,k} = \hat{p}^{n,k}(0)$.

2.3 The Central Limit Theorem

The main result of the paper is the following asymptotic normality statement.

Theorem 1. *Under Assumption 1, for any fixed $k \in \mathbb{N}^*$ and $a \in \mathbb{R}^+$, the following convergence in distribution holds true:*

$$\sqrt{n}\big(\hat{p}^{n,k} - p\big) \underset{n\to+\infty}{\to} \mathcal{N}\big(0, -p^2\log(p)\big). \tag{4}$$

Notice that the asymptotic variance does not depend on k. As a consequence of this result, one can define asymptotic Gaussian confidence intervals, for one realization of the algorithm and $n \to +\infty$. However, the speed of convergence is not known and may depend on the estimated probability p, and on the parameter k.

Thanks to Theorem 1, we can study the cost of the use of one realization of the AMS algorithm to obtain a given accuracy when $n \to +\infty$. In [3], the cost was analyzed when using a sample of M independent realizations of the algorithm, giving an empirical estimator, and the analysis was based on an asymptotic analysis of the variance in the large n limit.

Let ε be some fixed tolerance error, and $\alpha > 0$. Denote r_α such that $\mathbb{P}(Z \in [-r_\alpha, r_\alpha]) = 1 - \alpha$, where Z is a standard Gaussian random variable.

Then for n large, an asymptotic confidence interval with level $1-\alpha$, centered around p, is $[p - \frac{r_\alpha \sqrt{-p^2 \log(p)}}{\sqrt{n}}, p + \frac{r_\alpha \sqrt{-p^2 \log(p)}}{\sqrt{n}}]$. Then the ε-error criterion $|\hat{p}^{n,k} - p| \leq \varepsilon$ is achieved for n of size $\frac{-p^2 \log(p) r_\alpha^2}{\varepsilon^2}$.

However, on average one realization of the AMS algorithm requires a number of steps of the order $-n\log(p)/k$, with k random variables sampled at each iteration (see [3]). Another source of cost is the sorting of the replicas at initialization, and the insertion at each iteration of the k new sampled replicas in the sorted ensemble of the non-resampled ones. Thus the cost to achieve an accuracy of size ε is in the large n regime of size $n\log(n)\big(-p^2\log(p)\big)$, which does not depend on k.

This cost can be compared with the one when using a pure Monte-Carlo approximation, with an ensemble of non-interacting replicas of size n: thanks to the Central Limit Theorem, the tolerance criterion error ε is satisfied for n of size $\frac{-p(1-p) r_\alpha^2}{\varepsilon^2}$. Despite the $\log(n)$ factor in the AMS case, the performance is improved since $p^2 \log(p) = \mathrm{o}(p)$ when $p \to 0$.

Remark 1. In [3], the authors are able to analyze the effect of the change of k on the asymptotic variance of the estimator. Here, we do not observe significant differences when k changes, theoretically and numerically.

3 Proof of the Central Limit Theorem

The proof is divided into the following steps. First, thanks to Assumption 1, we explain why, in order to theoretically study the statistical behavior of the algorithm, it is sufficient to study to the case when X is distributed according to the exponential law with parameter 1: $\mathbb{P}(X > z) = \exp(-z)$ for any $z > 0$. The second step is the introduction of the characteristic function of $\log(\hat{p}^{n,k}(x))$; then, following the definition of the algorithm, we prove that it is solution of a functional equation with respect to x, which can be transformed into a linear ODE of order k. Finally, we study the solution of this ODE in the limit $n \to +\infty$.

3.1 Reduction to the exponential case

We first recall arguments from [3] which prove that it is sufficient to study the statistical behavior of the algorithm 1 and of the estimator 2 in a special case (Assumption 2 below); the more general result, Theorem 1 (valid under Assumption 1), is deduced from that special case.

It is sufficient to study the case when the random variable X is exponentially distributed with parameter 1. This observation is based on a change of variable with the following function:

$$\Lambda(x) = -\log\big(1 - F(x)\big). \tag{5}$$

It is well-known that $F(X)$ is uniformly distributed on $(0,1)$ (thanks to the continuity Assumption 1), and thus $\Lambda(X)$ is exponentially distributed with parameter 1. Thanks to Corollary 3.4 in [3], this property has the following consequence for the study of the AMS algorithm: the law of the estimator $\hat{p}^{n,k}$ is equal to the law of $\hat{q}^{n,k}$, which is the estimator defined, with (2), using the same values of the parameters n and k, but with two differences. First, the law of the underlying random variable is the exponential distribution with parameter 1; second, the stopping level a is replaced with $\Lambda(a)$, where Λ is defined by (5). Note the following consistency: $\mathbb{E}\left[\hat{q}^{n,k}\right] = \exp\left(-\Lambda(a)\right) = 1 - F(a) = p$ (by the unbiasedness result of [3]).

Since the arguments are intricate, we do not repeat them here and we refer the interested reader to [3]; from now on, we thus assume the following.

Assumption 2 *Assume that X is exponentially distributed with parameter* 1*: we denote $\mathscr{L}(X) = \mathscr{E}(1)$.*

When Assumption 2 is satisfied, the analysis is simpler and the rest of the paper is devoted to the proof of the following Proposition 1.

Proposition 1. *Under Assumption 2, the following convergence in distribution holds true:*

$$\sqrt{n}\left(\hat{p}^{n,k} - p\right) \underset{n\to+\infty}{\to} \mathscr{N}\left(0, a\exp(-2a)\right). \tag{6}$$

We emphasize again that even if the exponential case appears as a specific example (Assumption 2 obviously implies Assumption 1), giving a detailed proof of Proposition 1 is sufficient, thanks to Corollary 3.4 in [3], to obtain our main general result Theorem 1. Since the exponential case is more convenient for the computations below, in the sequel we work under Assumption 2. Moreover, we abuse notation: we use the general notations from Section 2, even under Assumption 2.

The following notations will be useful:

- $f(z) = \exp(-z)\mathbb{1}_{z>0}$ (resp. $F(z) = \left(1 - \exp(-z)\right)\mathbb{1}_{z>0}$) is the density (resp. the cumulative distribution function) of the exponential law $\mathscr{E}(1)$ with parameter 1.
- $f_{n,k}(z) = k\binom{n}{k}F(z)^{k-1}f(z)\left(1-F(z)\right)^{n-k}$ is the density of the k-th order statistics $X_{(k)}$ of a sample $(X_1,\dots,X_n)$, where the X_i are independent and exponentially distributed, with parameter 1.

Finally, in order to deal with the conditional distributions $\mathscr{L}(X|X>x)$ (which thanks to Assumption 2 is a shifted exponential distribution $x+\mathscr{E}(1)$) in the algorithm, we set for any $x \geq 0$ and any $y \geq 0$

$$\begin{aligned} f(y;x) = f(y-x), \quad F(y;x) = F(y-x), \\ f_{n,k}(y;x) = f_{n,k}(y-x), \\ F_{n,k}(y) = \int_{-\infty}^{y} f_{n,k}(z)dz, \quad F_{n,k}(y;x) = F_{n,k}(y-x). \end{aligned} \tag{7}$$

Straightforward computations (see also [3]) yield the following useful formulae:

$$\begin{cases} \frac{d}{dx} f_{n,1}(y;x) = n f_{n,1}(y;x). \\ \text{for } k \in \{2,\dots,n-1\},\ \frac{d}{dx} f_{n,k}(y;x) = (n-k+1)\left(f_{n,k}(y;x) - f_{n,k-1}(y;x)\right). \end{cases} \tag{8}$$

3.2 Proof of the Proposition 1

The first important idea is to prove Proposition 1 for all possible initial conditions $x \in [0,a]$, even if the value of interest is $x = 0$: in fact we prove the convergence

$$\sqrt{n}\big(\hat{p}^{n,k}(x) - p(x)\big) \underset{n\to+\infty}{\to} \mathscr{N}\big(0, (a-x)\exp(-2(a-x))\big). \tag{9}$$

A natural idea is to introduce the characteristic function of $\hat{p}^{n,k}(x)$, and to follow the strategy developed in [3]. Nevertheless, we are not able to derive a useful functional equation with respect to the x variable. The strategy we adopt is to study the asymptotic normality of the logarithm $\log(\hat{p}^{n,k}(x))$ of the estimator, and to use a particular case of the delta-method (see for instance [15], Section 3): if for a sequence of real random variables $(\theta_n)_{n\in\mathbb{N}}$ and a real number $\theta \in \mathbb{R}$ one has $\sqrt{n}\big(\theta_n - \theta\big) \underset{n\to\infty}{\to} \mathscr{N}(0,\sigma^2)$, then $\sqrt{n}\big(\exp(\theta_n) - \exp(\theta)\big) \underset{n\to\infty}{\to} \mathscr{N}\big(0, \exp(2\theta)\sigma^2\big)$, where convergence is in distribution.

We thus introduce for any $t \in \mathbb{R}$ and any $0 \le x \le a$

$$\phi_{n,k}(t,x) := \mathbb{E}\Big[\exp\Big(it\sqrt{n}\big(\log(\hat{p}^{n,k}(x)) - \log(P(x))\big)\Big)\Big]. \tag{10}$$

We also introduce an additional auxiliary function (using $P(x) = \exp(x-a)$)

$$\chi_{n,k}(t,x) := \mathbb{E}\Big[\exp\Big(it\sqrt{n}\hat{p}^{n,k}(x)\Big)\Big] = \exp\big(it\sqrt{n}(x-a)\big)\phi_{n,k}(t,x), \tag{11}$$

for which Lemma 1 states a functional equation, with respect to the variable $x \in [0,a]$. By Lévy's Theorem, Proposition 1 is a straightforward consequence (choosing $x = 0$) of Proposition 2 below.

Proposition 2. *For any $k \in \mathbb{N}^*$, any $0 \le x \le a$ and any $t \in \mathbb{R}$*

$$\phi_{n,k}(t,x) \underset{n\to+\infty}{\to} \exp\left(\frac{t^2(x-a)}{2}\right). \tag{12}$$

The rest of this section is devoted to the statement and the proof of four lemmas, and finally to the proof of Proposition 2.

Lemma 1 (Functional Equation). *For any $n \in \mathbb{N}$ and any $k \in \{1,\dots,n-1\}$, and for any $t \in \mathbb{R}$, the function $x \mapsto \chi_{n,k}(t,x)$ is solution of the following functional equation (with unknown χ): for any $0 \le x \le a$*

$$\chi(t,x) = e^{it\sqrt{n}\log(1-\frac{k}{n})} \int_x^a \chi(t,y) f_{n,k}(y;x)\,\mathrm{d}y \tag{13}$$

$$+ \sum_{l=0}^{k-1} e^{it\sqrt{n}\log(1-\frac{l}{n})} \mathbb{P}(S(x)^n_{(l)} < a \leq S(x)^n_{(l+1)}), \tag{14}$$

where $(S(x)^n_j)_{1\leq j\leq n}$ are iid with law $\mathscr{L}(X|X>x)$ and where $S(x)^n_{(l)}$ is the l-th order statistics of this sample (with convention $S(x)^n_{(0)} = x$).

Proof. The idea (like in the proof of Proposition 4.2 in [3]) is to decompose the expectation according to the value of the first level $Z^1 = X^0_{(k)}$. On the event $\{Z^1 > a\} = \{J^{n,k}(x) = 0\}$, the algorithm stops and $\hat{p}^{n,k}(x) = \frac{n-l}{n}$ for the unique $l \in \{0,\ldots,k-1\}$ such that $S(x)^n_{(l)} < a \leq S(x)^n_{(l+1)}$. Thus

$$\mathbb{E}[e^{it\sqrt{n}\log(\hat{p}^{n,k}(x))}\mathbb{1}_{J^{n,k}(x)=0}] = \sum_{l=0}^{k-1} e^{it\sqrt{n}\log(1-\frac{l}{n})} \mathbb{P}(S(x)^n_{(l)} < a \leq S(x)^n_{(l+1)}). \tag{15}$$

If $Z^1 < a$, for the next iteration the algorithm restarts from Z^1, and

$$\begin{aligned}
&\mathbb{E}[e^{it\sqrt{n}\log(\hat{p}^{n,k}(x))}\mathbb{1}_{J^{n,k}(x)>0}] \\
&= \mathbb{E}\left[e^{it\sqrt{n}\log(1-\frac{k}{n})}\mathbb{E}[e^{it\sqrt{n}\log\left(C^{n,k}(x)(1-\frac{k}{n})^{J^{n,k}(x)-1}\right)}|Z^1]\mathbb{1}_{Z^1<a}\right] \\
&= e^{it\sqrt{n}\log(1-\frac{k}{n})}\mathbb{E}\left[\mathbb{E}[e^{it\sqrt{n}\log\left(\hat{p}^{n,k}(Z^1)\right)}|Z^1]\mathbb{1}_{Z^1<a}\right] \\
&= e^{it\sqrt{n}\log(1-\frac{k}{n})}\mathbb{E}\left[\chi_{n,k}(t,Z^1)\mathbb{1}_{Z^1<a}\right] \\
&= e^{it\sqrt{n}\log(1-\frac{k}{n})}\int_x^a \chi_{n,k}(t,y) f_{n,k}(y;x)\,\mathrm{d}y.
\end{aligned} \tag{16}$$

Then (13) follows from (15), (16) and the definition (11) of $\chi_{n,k}$. □

We exploit the functional equation (13) for $x \mapsto \chi_{n,k}(t,x)$, to prove that this function is solution of a *Linear* Ordinary Differential Equation (ODE).

Lemma 2 (ODE). *Let n and $k \in \{1,\ldots,n-2\}$ be fixed. There exist real numbers $\mu^{n,k}$ and $(r^{n,k}_m)_{0\leq m\leq k-1}$, depending only on n and k, such that for all $t \in \mathbb{R}$, the function $x \mapsto \chi_{n,k}(t,x)$ satisfy the following Linear Ordinary Differential Equation (ODE) of order k: for $x \in [0,a]$*

$$\frac{d^k}{dx^k}\chi_{n,k}(t,x) = e^{it\sqrt{n}\log(1-\frac{k}{n})}\mu^{n,k}\chi_{n,k}(t,x) + \sum_{m=0}^{k-1} r^{n,k}_m \frac{d^m}{dx^m}\chi_{n,k}(t,x). \tag{17}$$

The coefficients $\mu^{n,k}$ and $(r^{n,k}_m)_{0\leq m\leq k-1}$ satisfy the following properties:

$$\mu^{n,k} = (-1)^k n \dots (n-k+1)$$
$$\lambda^k - \sum_{m=0}^{k-1} r_m^{n,k} \lambda^m = (\lambda - n) \dots (\lambda - n + k - 1) \quad \textit{for all } \lambda \in \mathbb{R}. \tag{18}$$

Observe that the ODE (17) is linear and that the coefficients are constant (with respect to the variable $x \in [0,a]$, for fixed parameters n,k and t). This nice property is the main reason why we consider the function $\chi_{n,k}$ (given by (11)) instead of $\phi_{n,k}$ (given by (10)); moreover it is also the reason why we study the characteristic function of $\log(\hat{p}^{n,k}(x))$, instead of the one of $\hat{p}^{n,k}(x)$.

Proof. The proof follows the same lines as Proposition 6.4 in [3]. We introduce

$$\Theta_{n,k}(t,x) := \sum_{l=0}^{k-1} e^{it\sqrt{n}\log(1-\frac{l}{n})} \mathbb{P}(S(x)^n_{(l)} < a \le S(x)^n_{(l+1)}).$$

Then by recursion, using the second line in (8), for $0 \le l \le k-1$ and for any $x \le a$ and $t \in \mathbb{R}$

$$\frac{d^l}{dx^l}\left(\chi_{n,k}(t,x) - \Theta_{n,k}(t,x)\right) = \mu_l^{n,k} e^{it\sqrt{n}\log(1-\frac{k}{n})} \int_x^a \chi_{n,k}(t,y) f_{n,k-l}(y;x)\,\mathrm{d}y + \sum_{m=0}^{l-1} r_{m,l}^{n,k} \frac{d^m}{dx^m}\left(\chi_{n,k}(t,x) - \Theta_{n,k}(t,x)\right), \tag{19}$$

with the associated recursion

$$\mu_0^{n,k} = 1, \mu_{l+1}^{n,k} = -(n-k+l+1)\mu_l^{n,k};$$
$$\begin{cases} r_{0,l+1}^{n,k} = -(n-k+l+1) r_{0,l}^{n,k}, \quad \text{if } l > 0, \\ r_{m,l+1}^{n,k} = r_{m-1,l}^{n,k} - (n-k+l+1) r_{m,l}^{n,k}, \quad 1 \le m \le l, \\ r_{l,l}^{n,k} = -1. \end{cases} \tag{20}$$

Using (19) for $l = k-1$ and the first line of (8), one eventually obtains, by differentiation, an ODE of order k:

$$\frac{d^k}{dx^k}\left(\chi_{n,k}(t,x) - \Theta_{n,k}(t,x)\right) = \mu^{n,k} e^{it\sqrt{n}\log(1-\frac{k}{n})} \chi_{n,k}(t,x) + \sum_{m=0}^{k-1} r_m^{n,k} \frac{d^m}{dx^m}\left(\chi_{n,k}(t,x) - \Theta_{n,k}(t,x)\right), \tag{21}$$

with $\mu^{n,k} := \mu_k^{n,k}$ and $r_m^{n,k} := r_{m,k}^{n,k}$.

It is key to observe that the coefficients $\mu^{n,k}$ and $(r_m^{n,k})_{0 \le m \le k-1}$ are defined by the same recursion as in [3]. In particular, they do not depend on the parameter $t \in \mathbb{R}$. To see a proof of (18), we refer to Section 6.4 in [3].

It is clear that the polynomial equality in (18) is equivalent to the following identity: for all $j \in \{0, \dots, k-1\}$

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

Due to the definition of the cumulative distribution functions of order statistics (7), one easily checks that $\Theta_{n,k}(t,.)$ is a linear combination of the exponential functions $x\mapsto \exp(nx),\dots,\exp((n-k+1)x)$; therefore

$$\frac{d^k}{dx^k}\Theta_{n,k}(t,x)=\sum_{m=0}^{k-1} r_m^{n,k}\frac{d^m}{dx^m}\Theta_{n,k}(t,x).$$

Thus the terms depending on $\Theta_{n,k}$ in (21) cancel out, and thus (17) holds true.

□

The next steps are to give an explicit expression of the solution of (17) as a linear combination of exponential functions, and to study the coefficients and the modes in the asymptotic regime $n\to+\infty$. Since the ODE is of order k, in order to uniquely determine the solution, more information is required: we need to know the derivatives of order $0,1,\dots,k-1$ of $x\mapsto \chi_{n,k}(t,x)$ at some point. We choose the terminal point $x=a$ (notice that by the change of variable $x\mapsto a-x$ the ODE (17) can then be seen as an ODE with an initial condition). This is the content of Lemma 3 below.

Lemma 3 (Terminal condition). *For any fixed $k\in\{1,\dots,\}$ and any $t\in\mathbb{R}$, we have*

$$\begin{cases}\chi_{n,k}(t,a)=1\\ \frac{d^m}{dx^m}\chi_{n,k}(t,x)\Big|_{x=a}\underset{n\to\infty}{=}O(\frac{1}{\sqrt{n}})n^m \quad \text{if } m\in\{1,\dots,k-1\}.\end{cases}\tag{22}$$

Proof. The equality $\chi_{n,k}(t,a)=1$ is trivial, since $\hat{p}^{n,k}(a)=1$. Equations (19) and (21), immediately imply (by recursion) that for $1\leq m\leq k-1$

$$\frac{d^m}{dx^m}\chi_{n,k}(t,x)\Big|_{x=a}=\frac{d^m}{dx^m}\Theta_{n,k}(t,x)\Big|_{x=a}.$$

Introduce the following decomposition

$$\begin{aligned}\Theta_{n,k}(t,x)&=\sum_{l=0}^{k-1}(e^{it\sqrt{n}\log(1-\frac{l}{n})}\mathbb{P}(S(x)^n_{(l)}<a\leq S(x)^n_{(l+1)})\\ &=\sum_{l=0}^{k-1}\left(e^{it\sqrt{n}\log(1-\frac{l}{n})}-1\right)\left(F_{n,l}(a;x)-F_{n,l+1}(a;x)\right)\\ &\quad+\sum_{l=0}^{k-1}\mathbb{P}(S(x)^n_{(l)}<a\leq S(x)^n_{(l+1)})\\ &=:\Omega_{n,k}(t,x)+1-F_{n,k}(a;x),\end{aligned}$$

where $F_{n,l}$ denotes the cumulative distribution function of the l-th order statistics (with the convention $F_{n,0}(a;x)=1$ for $x\leq a$), see (7).

Thanks to (8) and a simple recursion on l, it is easy to prove that for any $0 \leq l \leq k$ and any $m \geq 1$

$$\frac{d^m}{dx^m} F_{n,l}(a;x)\Big|_{x=a} = \mathrm{O}(n^m); \tag{23}$$

this immediately yields

$$\frac{d^m}{dx^m} \Omega_{n,k}(t,x)\Big|_{x=a} \underset{n\to\infty}{=} O(\frac{1}{\sqrt{n}})n^m.$$

In fact, it is possible to prove a stronger result: if $1 \leq l \leq k$ and $0 \leq m < l$ then

$$\frac{d^m}{dx^m} F_{n,l}(a;x)\Big|_{x=a} = 0,$$

by recursion on l and using (8) recursively on m. We thus obtain for $1 \leq m \leq k-1$

$$\frac{d^m}{dx^m}\big(1 - F_{n,k}(a;x)\big)\Big|_{x=a} = 0.$$

This concludes the proof of Lemma 3. $\square$

The last result we require is given by Lemma 4.

Lemma 4 (Asymptotic expansion). *Let $k \in \{1,\dots,\}$ and $t \in \mathbb{R}$ be fixed. Then for n large enough, we have*

$$\chi_{n,k}(t,x) = \sum_{l=1}^{k} \eta_{n,k}^l(t) e^{\lambda_{n,k}^l(t)(x-a)}, \tag{24}$$

for complex coefficients satisfying:

$$\begin{cases} \lambda_{n,k}^1(t) \underset{n\to\infty}{=} it\sqrt{n} + \frac{t^2}{2} + o(1), \\ \eta_{n,k}^1(t) \underset{n\to\infty}{\to} 1; \end{cases} \tag{25}$$

and for $2 \leq l \leq k$

$$\begin{cases} \lambda_{n,k}^l(t) \underset{n\to\infty}{\sim} n(1 - e^{\frac{i2\pi(l-1)}{k}}), \\ \eta_{n,k}^l(t) \underset{n\to\infty}{\to} 0. \end{cases} \tag{26}$$

Proof. We denote by $(\lambda_{n,k}^l(t))_{1\leq l\leq k}$ the roots of the characteristic equation associated with the linear ODE with constant coefficient (17) (with unknown $\lambda \in \mathbb{C}$): thanks to (18)

$$\frac{(n-\lambda)\dots(n-k+1-\lambda)}{n\dots(n-k+1)} - e^{it\sqrt{n}\log(1-\frac{k}{n})} = 0$$

By the continuity property of the roots of a complex polynomial of degree k with respect to its coefficients, we have

$$\overline{\lambda}_{n,k}^{l}(t) := \frac{\lambda_{n,k}^{l}(t)}{n} \underset{n\to\infty}{\to} \overline{\lambda}_{\infty}^{l},$$

where $(\overline{\lambda}_{\infty}^{l}(t))_{1\leq l\leq k}$ are the roots of $(1-\overline{\lambda})^k = 1$: thus $\lambda_{n,k}^{1}(t) and \underset{n\to\infty}{=} \mathrm{o}(n)$,

$$\lambda_{n,k}^{l}(t) \underset{n\to\infty}{\sim} n(1-e^{\frac{i2\pi(l-1)}{k}}).$$

To study more precisely the asymptotic behavior of $\lambda_{n,k}^{1}(t)$, we postulate an ansatz

$$\lambda_{n,k}^{1}(t) \underset{n\to\infty}{=} c_t\sqrt{n} + d_t + o(1);$$

We then identify the coefficients $c_t = it$ and $d_t = t^2/2$ thanks to the expansions

$$\left(1-\frac{c_t}{\sqrt{n}}-\frac{d_t}{n}+o(1)\right)^k \underset{n\to\infty}{=} 1-\frac{c_t k}{\sqrt{n}}-\frac{d_t k-\binom{k}{2}c_t^2}{n}+\mathrm{o}\left(\frac{1}{n}\right)$$
$$e^{it\sqrt{n}\log(1-\frac{k}{n})} \underset{n\to\infty}{=} 1-\frac{itk}{\sqrt{n}}-\frac{t^2k^2}{2n}+\mathrm{o}\left(\frac{1}{n}\right).$$

In particular, for n large enough, $(\lambda_{n,k}^{l}(t))_{1\leq l\leq k}$ are pairwise distinct, and (24) follows.

Then the coefficients $(\eta_{n,k}^{l}(t))_{1\leq l\leq k}$ are solutions of the following linear system of equations of order k:

$$\begin{cases} \eta_{n,k}^{1}(t)+\ldots+\eta_{n,k}^{k}(t) = \chi_{n,k}(a), \\ \eta_{n,k}^{1}(t)\overline{\lambda}_{n,k}^{1}(t)+\ldots+\eta_{n,k}^{k}(t)\overline{\lambda}_{n,k}^{k}(t) = \frac{1}{n}\frac{d}{dx}\chi_{n,k}(t,a), \\ \ldots \\ \eta_{n,k}^{1}(t)\left(\overline{\lambda}_{n,k}^{1}(t)\right)^{k-1}+\ldots+\eta_{n,k}^{k}(t)\left(\overline{\lambda}_{n,k}^{k}(t)\right)^{k-1} = \frac{1}{n^{k-1}}\frac{d^{k-1}}{dx^{k-1}}\chi_{n,k}(t,a). \end{cases} \tag{27}$$

Using Cramer's rule, we express each $\eta_{n,k}^{l}(t)$ as a ratio of determinants (the denominator is a Vandermonde determinant and is non zero when n is large enough). For $l \in \{2,\ldots,k\}$, we have

$$\eta_{n,k}^{l}(t) = \frac{\det(M_{n,k}^{l}(t))}{V\left(\overline{\lambda}_{n,k}^{1}(t),\ldots,\overline{\lambda}_{n,k}^{k}(t)\right)} \underset{n\to+\infty}{\to} 0,$$

where the matrix

$$M_{n,k}^{l}(t) = \begin{pmatrix} 1 & 1 & \ldots & 1 & \ldots & 1 \\ \overline{\lambda}_{n,k}^{1}(t) & \overline{\lambda}_{n,k}^{2}(t) & \ldots & O(\frac{1}{\sqrt{n}}) & \ldots & \overline{\lambda}_{n,k}^{k}(t) \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \left(\overline{\lambda}_{n,k}^{1}(t)\right)^{k-1} & \left(\overline{\lambda}_{n,k}^{2}(t)\right)^{k-1} & \ldots & O(\frac{1}{\sqrt{n}}) & \ldots & \left(\overline{\lambda}_{n,k}^{k}(t)\right)^{k-1} \end{pmatrix}$$

is such that $\det(M^l_{n,k}(t)) \underset{n\to+\infty}{\to} 0$ (since $\overline{\lambda}^1_{n,k}(t) \to 0$), while the denominator is the Vandermonde determinant $V\left(\overline{\lambda}^1_{n,k}(t),\ldots,\overline{\lambda}^k_{n,k}(t)\right) \underset{n\to+\infty}{\to} V\left(\overline{\lambda}^1_{\infty}(t),\ldots,\overline{\lambda}^k_{\infty}(t)\right) \neq 0$.

Finally, $\eta^1_{n,k}(t) = 1 - \sum_{l=2}^k \eta^l_{n,k}(t) \underset{n\to+\infty}{\to} 1$. This concludes the proof of Lemma 4. □

We are now in position to prove Proposition 2. Indeed, recall that $\phi_{n,k}(t,x) = \exp\left(-it\sqrt{n}(x-a)\right)\chi_{n,k}(t,x)$ thanks to (10) and (11). Then taking the limit $n \to +\infty$ thanks to Lemma 4 gives the convergence of the characteristic function $\phi_{n,k}$.

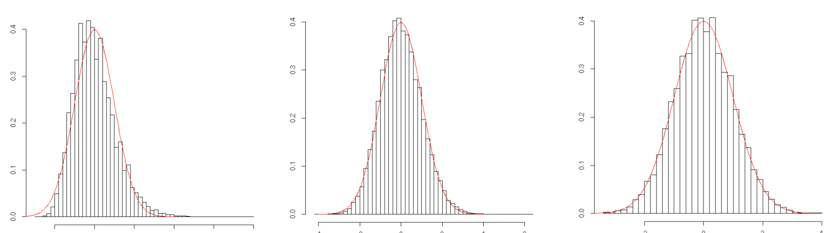
4 Numerical results

In this section, we provide numerical illustration of the Central Limit Theorem 1. We apply the algorithm with an exponentially distributed random variable with parameter 1 – this is justified by the discussion in Section 3.1.

In the simulations below, the estimated probability is $e^{-6}(\approx 2.48\ 10^{-3})$.

In Figure 1, we fix the value $k = 10$, and we show histograms for $n = 10^2, 10^3, 10^4$, with different values for the number M independent realizations of the algorithm, such that $nM = 10^8$ (we thus have empirical variance of the same order for all cases). In figure 1, we give the associated Q-Q plots, where the empirical quantiles of the sample are compared with the exact quantiles of the standard Gaussian random variable (after normalization).

Fig. 1 Histograms for $k = 10$ and $p = \exp(-6)$: $n = 10^2, 10^3, 10^4$ from left to right



In Figure 3, we show histograms for $M = 10^4$ independent realizations of the AMS algorithm with $n = 10^4$ and $k \in \{1, 10, 100\}$; we also provide associated Q-Q plots in figure 4.

From Figures 1 and 2, we observe that when n increases, the normality of the estimator is confirmed. Moreover, from Figures 3 and 4, no significant difference when k varies is observed.

Fig. 2 Q-Q plot for $k = 10$ and $p = \exp(-6)$: $n = 10^2, 10^3, 10^4$ from left to right

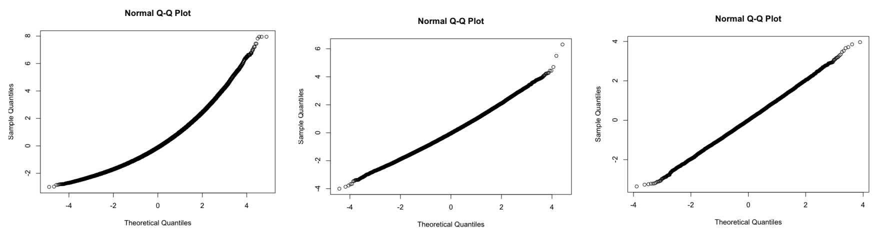


Fig. 3 Histograms for $n = 10^4$ and $p = \exp(-6)$: $k = 1, 10, 100$ from left to right

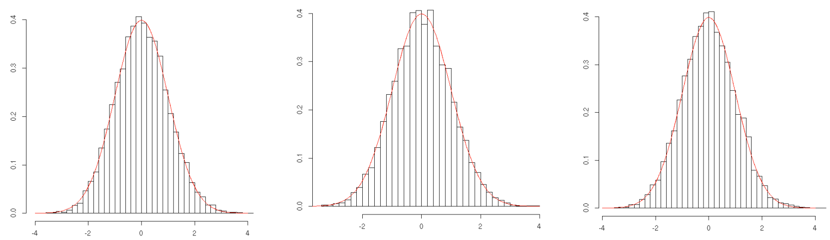
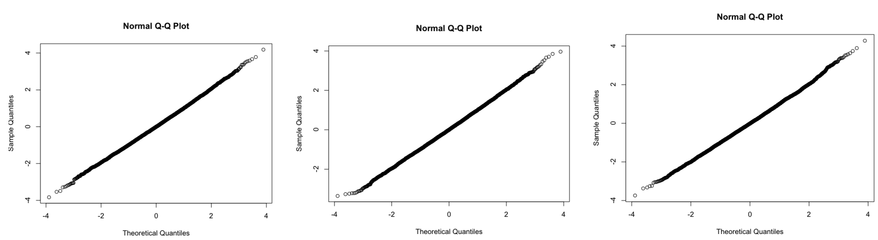


Fig. 4 Q-Q plot for $n = 10^4$ and $p = \exp(-6)$: $k = 1, 10, 100$ from left to right



Acknowledgements C.-E. B. would like to thank G. Samaey, T. Lelièvre and M. Rousset for the invitation to give a talk on the topic of this paper at the 11-th MCQMC Conference, in the special session on Mathematical aspects of Monte Carlo methods for molecular dynamics. We would also like to thank the referees for suggestions which improved the presentation of the paper.

References

1. S. Asmussen, P. W. Glynn. Stochastic simulation: Algorithms and analysis Springer, 2007.
2. S.K. Au and J.L. Beck. Estimation of Small Failure Probabilities in High Dimensions by Subset Simulation. *Journal of Probabilistic Engineering Mechanics*, Vol. 16, 263-277, 2001.

3. C.E. Bréhier T. Lelièvre, and M. Rousset. Analysis of Adaptive Multilevel Splitting Algorithms in an idealized case. *ESAIM Probability* & *Statistics*, to appear.
4. F. Cérou, P. Del Moral, T. Furon, and A. Guyader. Sequential Monte Carlo for rare event estimation. *Statistics and Computing*, 22(3):795–808, 2012.
5. F. Cérou and A. Guyader. Adaptive multilevel splitting for rare event analysis. *Stochastic Analysis and Applications*, 25(2):417–443, 2007.
6. 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.
7. F. Cérou, A. Guyader, T. Lelièvre, and D. Pommier. A multiple replica approach to simulate reactive trajectories. *Journal of Chemical Physics*, 134:054108, 2011.
8. F. Cérou, A. Guyader, P. Del Moral and F. Malrieu. Fluctuations of Adaptive Multilevel Splitting. *Arxiv e-preprints*, 2014.
9. P. Glasserman, P. Heidelberger, P. Shahabuddin and T. Zajic. Multilevel splitting for estimating rare event probabilities. *Operations Research*, 47(4), 585-600, 1999.
10. A. Guyader, N. Hengartner, and E. Matzner-Løber. Simulation and estimation of extreme quantiles and extreme probabilities. *Applied Mathematics* & *Optimization*, 64(2):171–196, 2011.
11. H. Kahn and T.E. Harris. Estimation of Particle Transmission by Random Sampling. *National Bureau of Standards, Appl. Math. Series*, 12: 27-30, 1951.
12. G. Rubino and B. Tuffin. Rare Event Simulation using Monte Carlo Methods. Wiley, 2009.
13. J. Skilling. Nested Sampling for General Bayesian Computation. *Bayesian Analysis*, 1 (4), 833-859, 2006.
14. E. Simonnet. Combinatorial analysis of the adaptive last particle method. *Statistics and Computing*, 2014.
15. A. W. van der Vaart. Asymptotic statistics. *Cambridge Series in Statistical and Probabilistic Mathematics*, 3, Cambridge University Press, Cambridge, 1998.