Applied Probability Trust (21 April 2009)

A TWO-STEP BRANCHING SPLITTING MODEL UNDER COST CONSTRAINT FOR RARE EVENT ANALYSIS

AGNÈS LAGNOUX-RENAUDIE,* *Université Toulouse 2*

Abstract

This paper deals with the splitting method first introduced in rare event analysis. In this technique, the sample paths are split into R multiple copies at various stages to speed up simulation. Given the cost, the optimization of the algorithm suggests to take all the transition probabilities equal; nevertheless, in practice, these quantities are unknown. In this paper, we present an algorithm in two steps that copes with that problem.

Keywords: splitting method, simulation, cost function, Laplace transform, Galton-Watson, branching processes, iterated functions, rare event

2000 Mathematics Subject Classification: Primary 65U05

Secondary 44A10;60J80

1. Introduction

The study of rare events is an important area in the analysis and prediction of major risks as earthquakes, floods, air collision risks, etc. Studying the major risks can be taken up by two main approaches which are the statistical analysis of collected data and the modelling of the processes leading to the accident. The statistical analysis of extreme values needs a long observation time since the very low probability of the events considered. The modelling approach consists first in formalizing the system

* Postal address: Institut de Mathématiques de Toulouse
Université Toulouse 2
31058 Toulouse France
URL : http://www.lsp.ups-tlse.fr/Fp/Lagnoux
Phone number : +33 (0)5 61 50 46 11
Fax : +33 (0)5 61 55 60 89

considered and then in using mathematical ([?] and [?]) or simulation tools to obtain some estimates.

Analytical and numerical approaches are useful, but may require many simplifying assumptions. On the other hand, Monte Carlo simulation is a practical alternative when the analysis calls for fewer simplifying assumptions. Nevertheless, obtaining accurate estimates of rare event probabilities, say about 10^{-9} to 10^{-12}, using traditional techniques require a huge amount of computing time.

Many techniques for reducing the number of trials in Monte Carlo simulation have been proposed, like importance sampling (see e.g. [?] and [?]) or trajectory splitting. In this technique, we suppose there exists some well identifiable intermediate system states that are visited much more often than the target states themselves and behave as gateway states to reach the rare event. Thus we consider a decreasing sequence of events L_i leading to the rare event L:

$$L = L_{M+1} \subset L_M \subset \cdots \subset L_1 \, .$$

Then $\mathbb{P}(L)$ is given by

$$\mathbb{P}(L) = \mathbb{P}(L|L_M)\mathbb{P}(L_M|L_{M-1}) \cdots \mathbb{P}(L_2|L_1)\mathbb{P}(L_1), \tag{1}$$

where on the right hand side, each conditioning event is "not rare". For the applications we have in mind, these conditional probabilities are in general not available explicitly. Instead we know how to make evolve the particles from level L_i to the next level L_{i+1} (e.g. Markovian behavior).

The principle of the algorithm is at first to run simultaneously several particles starting from the level L_i; after a while, some of them have evolved "badly", the other have evolved "well" i.e. have succeeded in reaching the threshold L_{i+1}. Then "bad" particles are moved to the position of the "good" ones and so on until L is reached. In such a way, the more promising particles are favoured; unfortunately that algorithm is hard to analyse directly because of the interaction introduced between particles and may be difficult to apply. Examples of this class of algorithms can be found in [?] with the "go with the winners" scheme, in [?, ?] in the context of the approximate counting and in [?, ?, ?] in a more general setting.

Nevertheless, all these algorithms lie on a common base, simpler to analyse and called branching splitting model. In this technique, we make a $\{0,1\}$ Bernoulli trial to check whether or not the set event L_1 has occured. In that case, we split this trial in R_1 Bernoulli subtrials, and for each of them we check again whether or not the event L_2 has occured. This procedure is repeated at each level, until L is reached. If an event level is not reached, neither is L, then we stop the current retrial. Using N independent replications of this procedure, we have then considered $NR_1 \cdots R_M$ trials, taking into account for example, that if we have failed to reach a level L_i at the i-th step, the $R_i \cdots R_M$ possible retrials have failed. Clearly the particles reproduce and evolve independently.

An unbiased estimator of $\mathbb{P}(L)$ is given by the quantity

$$\widehat{P} = \frac{N_L}{N \prod_{i=1}^{M} R_i}, \tag{2}$$

where N_L is the total number of trajectories having reached the set L. Considering that this algorithm is represented by N independent Galton-Watson branching processes $(Z_n)_n$, as done in [**?**], the variance of $\widehat{P}$ can be then derived and depends on the probability transitions and on the mean numbers (m_i) of particles successes at each level. Lead by the heuristic presented in [**?**, **?**], an optimal algorithm is derived by minimizing the variance of the estimator for a given budget (computational cost), defined as the expected number of trials generated during the simulation, where each trial is weighted by a cost function.

The optimization of the algorithm [?] suggests to take all the transition probabilities equal to a constant P_0 and the numbers of splitting equal to the inverse of this constant. We then deduce the number of thresholds M and finally N is given by the cost. This result is not surprising since it means that the branching processes are critical Galton-Watson processes. In other words, optimal values are chosen in such a way to balance the loss of variance from too little splitting and the exponential growth in computational effort from too much splitting.

Some pratical problems arises while applying the optimal algorithm in concrete models issued from reality. First, the optimal splitting number can be non integer. In [**?**], the author proposes three algorithms to face this problem. Then, for the applications we have in mind, the thresholds L_i are fixed but the conditional probabilities are unknown as said before (instead we know how to make evolve the particles from level L_i to the next level L_{i+1}). Moreover, we assume here that they lie in some compact $[a,b] \subset]0,1[$. This hypothesis is essential: otherwise, nothing can be done algorithmically. In practice, it is generally implicit but nothing is said about it.

We propose here an algorithm in two phases based on the branching splitting model: the first one is a learning phase in which we sample ρ_N particles. The algorithm proceeds as in the classical branching splitting method with splitting numbers $(R_i^0)_{i=1\cdots M}$ chosen arbitrarly at the beginning. In the second phase, we run $N-\rho_N$ particles that we make evolve as in the first phase but with splitting numbers estimators of the optimal splitting numbers $(R_i)_{i=1\cdots M}$; the estimators being obtained during the first learning phase and following the optimal rule given in [?]. Since the complexity of the formulas, we will simply lead an asymptotic study when the cost C goes to infinity. Assuming the transition probabilities lie in a compact implies that the cost by particle is bounded below and above which allows us to lead the survey when N goes to infinity. A precise analysis shows that we shall dedicate asymptotically $\mu_s C^{2/3}$ particles to the learning phase and $C/C_{opt} - \mu_s C^{2/3}$ to the second phase, where C_{opt} is a constant defined in Section 4.1 and μ_s is derived by the optimization of the algorithm; i.e. assuming that the number of particles generated during the learning phase behaves like $\mu_\alpha(C)C^{1-\alpha}$, we shall take $\alpha = 1/3$. Moreover, we note that N is linear in C, and so dedicating $\mu_s C^{2/3}$ particles to the first phase amounts to dedicate it $\lambda_s N^{2/3}$ particles, for some λ_s depending on μ_s.

The paper is organized as follows. Sections 2 recalls quickly the general settings and first results of the branching splitting model. We present, in Section 3, some useful analytical results. In section 4, we study the two-step branching splitting model and derive optimal parameters. Section 5 implements the two-step algorithm on an approximate Ornstein-Uhlenbeck process. Finally in Section 6, we conclude and discuss the merits of this approach and potential directions for further researches.

2. Previous results and general settings

2.1. Optimal branching splitting model

As said in the Introduction and following [?], we consider N independent Galton-Watson branching processes $(Z_n^{(i)})_{n\geq 0}$, $i = 1, \cdots, N$ where for each i, $Z_n^{(i)}$ is the number of particles derived from the i-th particle ($Z_0^{(i)}$=1) that have reached the level L_n. Then, letting R_i the sampling number at level i,

$$\widehat{P} := \frac{1}{N}\sum_{i=1}^{N}\widetilde{Q}_i, \quad \text{where} \quad \widetilde{Q}_i = \frac{Z_{M+1}^{(i)}}{R_1\cdots R_M}. \tag{3}$$

To lighten notation, we will consider only the case $N = 1$ in the following, i.e. we will consider the process $(Z_n)_{n\geq 0}$ with $Z_0 = 1$. We have the following relation

$$Z_{n+1} = \sum_{j=1}^{Z_n} X_n^{(j)} \tag{4}$$

where for each n, the random variables $(X_n^{(j)})_{j\geq 1}$ are i.i.d. with common law a binomial distribution with parameters (R_n, P_{n+1}) for $n \geq 1$ and a Bernoulli distribution with parameter P_1 for $n = 0$. The reader is referred to Harris [**?**], Lyons [**?**] and Athreya and Ney [**?**] for more details on Galton-Watson and branching processes.

Let introduce the following quantities

$$\begin{aligned}
r_0 &= 1, \qquad r_i = R_1\cdots R_i, \quad i = 1\cdots M,\\
p_0 &= 1, \qquad p_i = P_1\cdots P_i, \quad i = 1\cdots M+1,\\
m_0 &= P_1, \qquad m_i = P_{i+1}R_i, \quad i = 1\cdots M,
\end{aligned}$$

where r_i represent the weight of a praticle having reached L_i, p_i the probability to reach L_i starting from 0 and m_i the mean number of offsprings of an individual at i-th generation. Integrating N, the variance of $\widehat{P}$ is given by (see [?] for details)

$$\operatorname{var}(\widehat{P}) = \frac{\mathbb{P}(L)^2}{N}\sum_{i=0}^{M}\left(\frac{1}{p_{i+1}} - \frac{1}{p_i}\right)\frac{1}{r_i}. \tag{5}$$

In [?], the three steps minimization of the variance of $\mathbb{P}(L)$ for a given budget C defined as the mean number of particles generated during the simulation

$$C = N\sum_{i=0}^{M} r_i p_i \tag{6}$$

leads to the optimal parameters of the algorithm: first we derive the optimal N and the optimal splitting numbers R_i, then the optimal transition probabilities P_i and finally the optimal number M of thresholds. In the case where the transition probabilities are fixed, the first step of the optimization leads to

$$R_i = \sqrt{\frac{1}{P_i P_{i+1}}}\sqrt{\frac{1-P_{i+1}}{1-P_i}} \qquad \text{for } i = 1\cdots M \tag{7}$$

$$N = \frac{C\sqrt{1/P_1 - 1}}{\sum_{i=1}^{M+1}\sqrt{1/P_i - 1}} \tag{8}$$

Using these relations, the variance has the following expression

$$\mathrm{var}(\widehat{P}) = \frac{\mathbb{P}(L)^2}{N}\left(\frac{1}{P_1} - 1\right)\sum_{i=0}^{M} r_i p_i = \frac{\mathbb{P}(L)^2}{N^2}\left(\frac{1}{P_1} - 1\right)C \tag{9}$$

2.2. Settings

2.2.1. *The two-step branching splitting model* In this paper, the transition probabilities are considered as unknown but belong to some interval $[a,b] \subset]0,1[$. In practice, one can propose values for a and b by a specific knowledge of the process or by a previous learning simulation when the are unknown. Choose arbitrarly $M+1$ numbers P_i^0 in $[a,b]$, let

$$R_i^0 = \frac{1}{\sqrt{P_i^0 P_{i+1}^0}}\sqrt{\frac{1-P_{i+1}^0}{1-P_i^0}} \quad \text{for} \quad i = 1\cdots M$$

and introduce the following notation

$$r_0^0 = 1, \qquad r_i^0 = R_1^0 \cdots R_i^0, \quad i = 1\cdots M.$$

Remark 2.1. Moreover, the analytical complexity of the calculation prevents us to make a precise analysis and we are lead to derive only asymptotic results when the cost C goes to infinity. But the fact that the P_i's lie in the compact $[a,b] \subset]0,1[$ implies that the ratio C/N is bounded above and below allowing us to make the asymptotic study while N goes to infinity. Note also that the total number of particles is in reality a random variable: in practice, a given budget is fixed at the beginning and we sample particles until the budget is consumed; hence the randomness of N.

We propose here an algorithm in two phases: the first one is a learning phase in which the splitting numbers are $(R_i^0)_{i=1\cdots M}$ while the second one uses estimators of $(R_i)_{i=1\cdots M}$ obtained in the learning phase. More precisely, we have N particles that we split in two groups of ρ_N and $N-\rho_N$ particles respectively and the two-step branching splitting model proceeds as follows:

Phase 1: We sample ρ_N particles. The particles having reached L_1 are split into R_1^0 subtrials that we make evolve from L_1. The particles having reached L_2 are split into R_2^0 subtrials that we make evolve from $L_2 \cdots$ And so on until L is reached. From step i ($i = 1 \cdots M+1$), we get an estimator $\widehat{P}_i^{(1)}$ of P_i (fraction of successful particles starting from L_{i-1}). To exploit all the information we possess, we improve the algorithm replacing $\widehat{P}_i^{(1)}$ by $\widetilde{P}_i^{(1)} := a \vee \widehat{P}_i^{(1)} \wedge b$ during the simulation, since $P_i \in [a, b]$. This substitution have the convenient consequence to prevent us from any premature stop of the algorithm. Without this assumption, the possibility of the particle dying remains that can be controlled probabilistically (see Section 3). After a premature stop of the algorithm, one can start a new simulation. Nevertheless, one must take into account the cost induced by this first aborted phase since we work for a given fixed effort. Now let for all $i = 1 \cdots M$,

$$\widetilde{R}_i = \frac{1}{\sqrt{\widetilde{P}_i^{(1)} \widetilde{P}_{i+1}^{(1)}}} \sqrt{\frac{1-\widetilde{P}_{i+1}^{(1)}}{1-\widetilde{P}_i^{(1)}}},$$

they represent the splitting numbers of the second phase.

Phase 2: We sample $N-\rho_N$ particles. The particles having reached L_1 are split into $\widetilde{R}_1$ subtrials that we make evolve from L_1. The particles having reached L_2 are split into $\widetilde{R}_2$ subtrials that we make evolve from $L_2 \cdots$ And so on until L is reached. From step i ($i = 1 \cdots M+1$), we get an estimator $\widehat{P}_i^{(2)}$ of P_i that we change into $\widetilde{P}_i^{(2)} := a \vee \widehat{P}_i^{(2)} \wedge b$ during the simulation.

In the following, the truncated estimators will be denoted with a tilde $\sim$ and the others with a hat $\wedge$.

Assumptions:

1. A first analytical survey leads us to restrict ourselves to the case when

   - $\rho_N \underset{N\to+\infty}{\to} +\infty$,

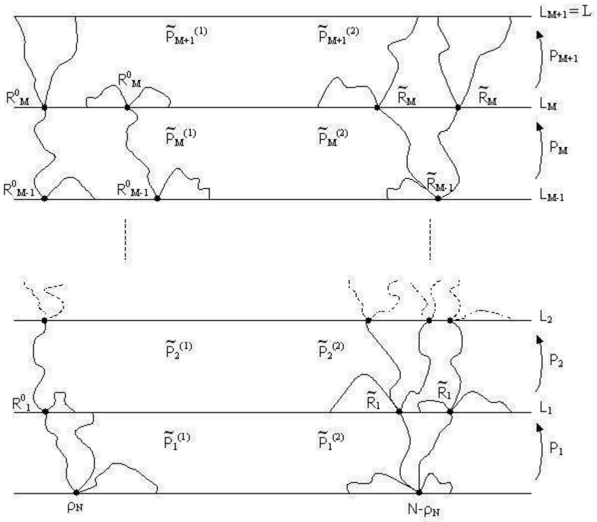

FIGURE 1: The two-step branching splitting model

- $N - \rho_N \underset{N\to+\infty}{\to} +\infty$,
- $\rho_N = o(N - \rho_N)$.

2. Moreover, as said in Remark 2.1, the calculation being too complex to make a precise analysis, we will lead an asymptotic study when the cost C (and so when N) goes to infinity. We aim at optimizing the algorithm in the parameters. After analysis, we prove that the optimum is attained when

$$\left(\frac{\rho_N}{N}\right)^2 \approx \frac{1}{\rho_N}, \qquad \text{while} \;\; N \to \infty$$

More precisely, to have a better understanding on what happens and to clarify the analysis, we restrict ourselves to the case where ρ_N is asymptotically of the form $\lambda N^{1-\alpha}$ with λ depending on N and we aim at optimizing the algorithm in α, N and λ. Then we prove that the optimum is attained for $\alpha = 1/3$: indeed, for $\alpha < 1/3$, $\lambda \underset{N\to\infty}{\longrightarrow} 0$, for $\alpha = 1/3$, $\lambda \underset{N\to\infty}{\longrightarrow}$ cst and for $\alpha > 1/3$, $\lambda \underset{N\to\infty}{\longrightarrow} \infty$.

Notation: Let

$$\mathcal{F}_{2i} = \sigma\left(\widetilde{P}_1^{(1)}, \cdots, \widetilde{P}_i^{(1)}, \widetilde{P}_1^{(2)}, \cdots, \widetilde{P}_i^{(2)}\right) \quad \text{and} \quad \mathcal{F}_{2i+1} = \mathcal{F}_{2i} \vee \sigma\left(\widetilde{P}_{i+1}^{(1)}\right)$$

and we note, in the following, $\mathbb{E}_k$ for $\mathbb{E}(\bullet|\mathcal{F}_k)$ and var_k for $\mathrm{var}(\bullet|\mathcal{F}_k)$.

2.2.2. *Estimators of the P_i's* To exploit all the information given during the simulation, we use both algorithms to estimate the transition probabilities: P_{i+1} is estimated by the fraction of successful particles in each generation in phase 1 and in phase 2. More precisely, for $i = 0 \cdots M$, P_{i+1} is estimated by

$$\widetilde{P}_{i+1} = \frac{\text{total nb of successes in } L_{i+1}}{\text{total nb of particles generated from } L_i} = \frac{\frac{\lambda}{N^\alpha} r_i^0 \widetilde{p}_{i+1}^{(1)} + (1 - \frac{\lambda}{N^\alpha})\widetilde{r}_i \widetilde{p}_{i+1}^{(2)}}{\frac{\lambda}{N^\alpha} r_i^0 \widetilde{p}_i^{(1)} + (1 - \frac{\lambda}{N^\alpha})\widetilde{r}_i \widetilde{p}_i^{(2)}}$$

with

$$\begin{cases} \widetilde{p}_0^{(1)} = 1, & \widetilde{p}_i^{(1)} = \widetilde{P}_1^{(1)} \cdots \widetilde{P}_i^{(1)} \text{ for } i = 1 \cdots M+1 \\ \widetilde{p}_0^{(2)} = 1, & \widetilde{p}_i^{(2)} = \widetilde{P}_1^{(2)} \cdots \widetilde{P}_i^{(2)} \text{ for } i = 1 \cdots M+1 \\ \widetilde{r}_0 = 1, & \widetilde{r}_i = \widetilde{R}_1 \cdots \widetilde{R}_i \text{ for } i = 1 \cdots M \end{cases}$$

It corresponds to the minimal variance estimator among the estimators linear combination of $\widetilde{P}_i^{(1)}$ and $\widetilde{P}_i^{(2)}$.

Finally, $\mathbb{P}(L)$ is estimated by the product of these fractions:

$$\widetilde{P} = \widetilde{P}_1 \cdots \widetilde{P}_{M+1} =: \widetilde{p}_{M+1}$$

and noting that

$$\widetilde{P}_{i+1} = \frac{\widetilde{p}_{i+1}^{(2)}}{\widetilde{p}_i^{(2)}} \left[1 + \frac{\lambda}{N^\alpha}\left(\frac{\widetilde{p}_{i+1}^{(1)} r_i^0}{\widetilde{p}_{i+1}^{(2)} \widetilde{r}_i} - 1\right)\right] \Big/ \left[1 + \frac{\lambda}{N^\alpha}\left(\frac{\widetilde{p}_i^{(1)} r_i^0}{\widetilde{p}_i^{(2)} \widetilde{r}_i} - 1\right)\right]$$

we have the following result

Proposition 2.1. $\mathbb{P}(L)$ *is estimated by*

$$\widetilde{P} = \widetilde{p}_{M+1}^{(2)} \frac{\prod_{i=1}^{M+1}\left\{1 + \frac{\lambda}{N^\alpha}\left(\frac{\widetilde{p}_i^{(1)} r_{i-1}^0}{\widetilde{p}_i^{(2)} \widetilde{r}_{i-1}} - 1\right)\right\}}{\prod_{i=1}^{M}\left\{1 + \frac{\lambda}{N^\alpha}\left(\frac{\widetilde{p}_i^{(1)} r_i^0}{\widetilde{p}_i^{(2)} \widetilde{r}_i} - 1\right)\right\}} \tag{10}$$

3. Mathematical tools

In this section, we present the technical tools used for the rest of the analysis: we give two results concerning the truncated estimators used during the algorithm. First a precise bound of the truncation probabilities of a sum of i.i.d. Bernoulli random variables is presented; then we state a result on the expectation of truncated estimators.

Lemma 3.1. *Let $\widehat{P}$ a random variable having the following expression*

$$\widehat{P} = \frac{1}{N}\sum_{i=1}^{N} Ber_i$$

where Ber_i are i.i.d. Bernoulli random variables with parameter P.

1. *First, letting $l(x) = \log\left[\left(\frac{1-x}{1-P}\right)^{1-x}\left(\frac{x}{P}\right)^{x}\right]$, we have*

$$\begin{aligned}\mathbb{P}(\widehat{P} \le a) &\le \exp\{-Nl(a)\}\\ \mathbb{P}(\widehat{P} \ge b) &\le \exp\{-Nl(b)\}\end{aligned}$$

2. *Let f, g two functions defined on $]0,1[$, $\mathcal{C}^2$ on $[a,b]$ and $\widetilde{P} = a \vee \widehat{P} \wedge b$. Then*

$$\mathbb{E}\left(f(\widetilde{P})\right) = f(P) + \frac{P(1-P)}{2N} f''(P) + o(\frac{1}{N})$$

As a consequence,

$$\begin{aligned}var\left(f(\widetilde{P})\right) &= \frac{P(1-P)}{N} f'(P)^2 + o(\frac{1}{N})\\ cov\left(f(\widetilde{P}), g(\widetilde{P})\right) &= \frac{P(1-P)}{N} f'(P)g'(P) + o(\frac{1}{N})\end{aligned}$$

The proof of this technical lemma is postponed to Appendix A. One can easily generalize these results to the functions of two variables by duplication.

4. Asymptotic optimal algorithm

4.1. Asymptotic expressions for the cost and the variance

In this section, we want to derive the optimal algorithm by minimizing the variance of the estimator for a given simulation cost. As mentioned in Remark 2.1, we will lead an asymptotic survey while N goes to infinity and derive an asymptotic optimal algorithm by minimizing the variance for a given budget. First, we give asymptotic expressions for the cost and the variance; then we deduce an asymptotic optimal algorithm and we end the section explaining how to proceed practically.

Thus, we first need to derive asymptotic expressions of the (average) simulation cost and the variance of the estimator. An accurate analysis of the optimization problem shows that we must do an asymptotic expansion at the second order for the learning phase (i.e. at the second order in ρ_N) while we simply need the first order for the second phase (i.e. at the first order in $N - \rho_N$ i.e. N). The proofs of the two following theorems being technical and complex are postponed to Appendix B and C.

Assuming that asymptotically, ρ_N as the form $\lambda N^{1-\alpha}$ as said previously, the (average) cost (which is defined as the mean number of particles generated during the simulation as done in Section 2) is given by

$$C = \lambda N^{1-\alpha} \sum_{i=0}^{M} R_0^i \mathbb{E}\left(\widetilde{p}_i^{(1)}\right) + (N - \lambda N^{1-\alpha}) \sum_{i=0}^{M} \mathbb{E}\left(\widetilde{r}_i \widetilde{p}_i^{(2)}\right) \tag{11}$$

Remember that we aim at deriving the expression of the cost as the given budget goes to infinity. We remark that

$$\frac{C}{N} = \frac{\lambda}{N^\alpha} \sum_{i=0}^{M} R_0^i \mathbb{E}(\widetilde{p}_i^{(1)}) + \left(1 - \frac{\lambda}{N^\alpha}\right) \sum_{i=0}^{M} \mathbb{E}(\widetilde{r}_i \widetilde{p}_i^{(2)})$$

and since $P_i \in [a, b] \subset]0, 1[$ for all $i = 1 \cdots M+1$, the right hand side is bounded above and below; as a consequence, we can verify that as announced in the Introduction,

$$N \to \infty \;\; \text{while} \;\; C \to \infty$$

and C is linear in N. Now using repeatedly Lemma 3.1, we derive

Theorem 4.1.

$$C = N \left[C_{opt} + \frac{B_1}{\lambda N^{1-\alpha}} + B_2 \frac{\lambda}{N^\alpha} + o\left(\frac{1}{N^\alpha}\right) + o\left(\frac{1}{N^{1-\alpha}}\right) \right]$$

with

$$\begin{cases} B_1 = \frac{1}{2}\sum_{i=1}^{M} r_i p_i \left[\frac{3/4}{P_1(1-P_1)} - 2 + 2\sum_{k=1}^{i-1} \frac{1-P_{k+1}}{r_k^0 p_{k+1}} + \frac{3/4-P_{i+1}}{r_i^0 p_{i+1}(1-P_{i+1})}\right] \\ B_2 = \sum_{i=1}^{M} p_i (r_i^0 - r_i) \end{cases}$$

and C_{opt} the (average) particle-cost generated in the optimal model, given by

$$\sum_{i=0}^{M} r_i p_i$$

Using the notation introduced previously, the variance of the estimator is given by

Theorem 4.2.

$$var(\widetilde{P}) = \frac{\mathbb{P}(L)^2}{N}\left[V_{opt} + \frac{1}{\lambda N^{1-\alpha}} A_1 + \frac{\lambda}{N^{\alpha}} A_2 + \frac{\lambda^2}{N^{2\alpha}} A_3 + o\left(\frac{1}{N^{1-\alpha}}\right) + o\left(\frac{1}{N^{2\alpha}}\right)\right]$$

with

$$\begin{cases} A_1 = \frac{1}{2}\sum_{i=1}^{M} \left[\frac{P_{i+1}-1/4}{r_i^0 r_i p_{i+1}^2} - \frac{1}{4}\frac{r_i p_i}{P_1^2}\right] \\ A_2 = \left(\frac{1}{P_1} - 1\right)\sum_{i=1}^{M} p_i (r_i - r_i^0) \\ A_3 = \left(\frac{1}{P_1} - 1\right)\sum_{i=1}^{M} \frac{p_i}{r_i}(r_i^0 - r_i)^2 \end{cases}$$

and V_{opt} the renormalized particle-variance of the optimal model, given by

$$\sum_{i=0}^{M} \left(\frac{1}{p_{i+1}} - \frac{1}{p_i}\right)\frac{1}{r_i} = \left(\frac{1}{P_1} - 1\right) C_{opt}, \tag{12}$$

coming from equation (9).

4.2. Optimization of the algorithm

We remind the reader that our goal is to optimize the algorithm in the parameters N, α and λ. We first prove the following lemma:

Lemma 4.1. *Let F and G two real-valued functions. Suppose that*

$$\begin{aligned} F(\lambda, N) &\underset{N\to\infty}{\sim} \frac{1}{N} + \frac{f_1(\lambda)}{N^{1+\alpha}} + \frac{f_2(\lambda)}{N^{1+2\alpha}} + \frac{f_3(\lambda)}{N^{2-\alpha}}, \\ G(\lambda, N) &\underset{N\to\infty}{\sim} N - N^{1-\alpha} f_1(\lambda) + N^{\alpha} f_4(\lambda). \end{aligned}$$

Then minimizing F for a fixed $G = G_f$ asymptotically amounts to minimize the function of λ

- $f_3 + f_4$ *for* $\alpha > \frac{1}{3}$,
- $f_2 - \alpha(f_1)^2$ *for* $\alpha < \frac{1}{3}$,
- $f_2 + f_3 + f_4 - (f_1)^2$ *for* $\alpha = \frac{1}{3}$,

N *being given by the equation* $G = G_f$.

The proof of this lemma is postponed to Appendix D. As a consequence, we get the following proposition:

Proposition 4.1. *Minimizing the variance for a given cost leads to take* $\alpha = \frac{1}{3}$ *and to the solution*

$$\lambda_s = \left[\frac{C_{opt}\left(A_1 + B_1\left(\frac{1}{P_1} - 1\right)\right)}{2\left(A_2 B_2 + A_3 C_{opt}\right)}\right]^{\frac{1}{3}},$$

N *being given by the cost.*

Proof. From equation (9) and the expressions of A_2 and B_2, the following relation

$$A_2 C_{opt} + B_2 V_{opt} = 0 \tag{13}$$

is clearly satisfied. As a consequence, the variance and the cost (after a renormalization by V_{opt} and C_{opt} respectively) have the form of functions F and G in Lemma 4.1 which allows us to state that minimizing the variance for a given cost C amounts to minimizing the functions

- for $\alpha > \frac{1}{3}$, $\frac{1}{\lambda}\left[\frac{A_1}{V_{opt}} + \frac{B_1}{\lambda C_{opt}}\right]$ which leads to $\lambda_s = +\infty$.
- for $\alpha < \frac{1}{3}$, $\lambda^2\left[\frac{A_3}{V_{opt}} + \alpha\frac{A_2 B_2}{V_{opt} C_{opt}}\right]$ which leads to $\lambda_s = 0$.
- for $\alpha = \frac{1}{3}$, $\frac{\frac{A_1}{\lambda} + A_3\lambda^2}{V_{opt}} + \frac{B_1}{\lambda C_{opt}} + \lambda^2\frac{A_2 B_2}{V_{opt} C_{opt}}$ which leads to $\lambda_s = \left[\frac{C_{opt}\left(A_1 + B_1\left(\frac{1}{P_1} - 1\right)\right)}{2(A_2 B_2 + A_3 C_{opt})}\right]^{\frac{1}{3}}$.

Finally one must take

$$\alpha = \frac{1}{3},$$

N is given by the cost and the solution of the optimization problem by

$$\lambda_s = \left[\frac{C_{opt}\left(A_1 + B_1\left(\frac{1}{P_1} - 1\right)\right)}{2\left(A_2 B_2 + A_3 C_{opt}\right)}\right]^{\frac{1}{3}}$$

Remark 4.1. We insist on the fact that equation $A_2 C_{opt} + B_2 V_{opt} = 0$, allowing us to apply Lemma 4.1, directly comes from the particular choice of the R_i's, as the solutions of a particular optimization problem and this leads to $\alpha = \frac{1}{3}$. Otherwise, we certaintly would not have had this relation between the constant A_2 and B_2 and the optimal α would simply have been $\frac{1}{2}$, as one can expect to a priori. This emphasizes on the fact that it is really worth to choose the transition probabilities as close as possible to the optimals.

Remark 4.2. If we take arbitrarly the optimal R_i for the learning phase, we are lead to the solution

$$\lambda_s = +\infty$$

that matches the intuition: if the optimal splitting numbers are used from the start then introducing an extra Monte Carlo stage cannot be advantageous.

4.3. Guidelines to proceed practically

In practice, we are given a fixed budget to consume during the simulation which amounts to fix the total number N of particles generated during the simulation. We will sample $\rho_N = \lambda_s N^{1-\alpha}$ during the learning phase and $N - \rho_N$ to the second. The optimization analysis suggets us to take

$$\alpha = \frac{1}{3} \quad \text{and} \quad \lambda_s = \left[\frac{C_{opt} \left(A_1 + B_1 \left(\frac{1}{P_1} - 1 \right) \right)}{2 \left(A_2 B_2 + A_3 C_{opt} \right)} \right]^{\frac{1}{3}}.$$

Recall that the transition probabilities P_i are unknown and thus we can not evaluate explicitly λ_s. Nevertheless, these conditional probabilities are bounded above and below; thus we can determine a lower bound $\lambda_s(\min)$ (but also an upper bound $\lambda_s(\max)$) of λ_s. As a consequence, we will proceed as follows:

- First phase-*learning phase*:

  We first sample $\lambda_s(\min) N^{2/3}$ particles, proceeding as in the classical branching splitting method with splitting numbers $(R_i^0)_{i=1\cdots M}$ that have been chosen arbitrarily at the beginning. Then we estimate the transition probabilities, the optimal splitting numbers and so the optimal λ_s by λ_{est}. We finally continue to sample particles until the total number of trials generated during the first phase attained $\lambda_{est} N^{2/3}$.

- Second phase:

  We run $N - \rho_N$ particles that we make evolve as in the first phase but here we use estimators of $(R_i)_{i=1 \cdots M}$ obtained during the first learning phase and following the optimal rule given in [?], as explained previously.

5. Approximate Ornstein-Uhlenbeck process

We aim at study the process called *approximate Ornstein-Uhlenbeck process* and governed by the following stochastic differential equation:

$$dX_t = -\mu(X_t)X_t dt + \sigma(X_t)dW_t \tag{14}$$

where the unknown drift is such that

$$0 < \mu^1 \leq \mu(x) \leq \mu^2 < \infty.$$

In the numerical application, we will take the drift constant on disjoint intervals. Reminber that the general Ornstein-Uhlenbeck process is governed by the following stochastic differential equation:

$$dX_t = -\mu X_t dt + \sigma dW_t \tag{15}$$

The Ornstein-Uhlenbeck process is recurrent and excursions above large values are rare; which is also the case for the approximate Ornstein-Uhlenbeck process. Here, we want to estimate the probabilities that the approximate Ornstein-Uhlenbeck process reaches some high levels starting from $x > 0$ and before returning to 0. When the drift is known, one can easily determine these probabilities. But with an unknown drift, its estimation is really expensive in terms of simulation. Here, since we just want to estimate the levels of excursions, there is no need to estimate μ and the two-step branching splitting algorithm appears to be a promising and efficient way to solve the problem.

Before leading the numerical application, recall, first, the general settings and results on the general Ornstein-Uhlenbeck process.

5.1. Analytical study

We recall, in that section, some results on the standard Ornstein-Uhlenbeck process, governed by the stochastic differential equation:

$$dX_t = -X_t dt + \sqrt{2} dW_t \tag{16}$$

(i.e. drift μ equal to 1, variance σ^2 to 2). The results for the general Ornstein-Uhlenbeck process governed by (15) can be easily deduced by a change of variables.

The speed measure m is given by

$$m(dx) = \rho(x)dx = e^{-\mu x^2/\sigma^2} dx$$

and the infinitesimal generator $\mathcal{L}$ by

$$\mathcal{L}(f)(x) = f''(x) - \mu x f'(x) = f''(x) + \frac{\sigma^2}{2} \frac{\rho'(x)}{\rho(x)} f'(x)$$

Let L_0, x and L such as $L_0 < x < L$. Suppose $X_0 = x$, note $H(x) = \int_{L_0}^{x} \frac{dy}{\rho(y)}$ and $T_{L_0,L}$ the first leaving time of $[L_0, L]$:

$$T_{L_0,L} = \inf\{t \geq 0 : X_t \notin [L_0, L]\}$$

One can easily verify that $T_{L_0,L}$ is a stopping time and noting that H solves $\mathcal{L}(f) = 0$, we have

$$\begin{cases} \mathbb{P}_x(X_{T_{L_0,L}} = L) = \frac{H(x)}{H(L)} \\ \mathbb{P}_x(X_{T_{L_0,L}} = L_0) = \frac{H(L)-H(x)}{H(L)} \end{cases}$$

$$\mathbb{E}_x(T_{L_0,L}) = M_{L_0,L}(x) = -\int_{L_0}^{x} \frac{R(t)}{\rho(t)} dt + \frac{H(x)}{H(L)} \int_{L_0}^{L} \frac{R(t)}{\rho(t)} dt \tag{17}$$

How to position the thresholds to have constant transitions The optimization of the algorithm leads to take the transition probabilities equal i.e. if the thresholds are noted L_i, we want to solve in L_{i+1} once L_i having be fixed

$$\mathbb{P}_{L_i}(X_{T_{L_0,L_{i+1}}} = L_{i+1}) = \frac{H(L_i)}{H(L_{i+1})} = \theta$$

for some θ. Then, for large values of i, we aim at having

$$\begin{aligned}\mathbb{P}_{L_i}(X_{T_{L_0,L_{i+1}}} = L_{i+1}) = \theta \quad &\Leftrightarrow \quad H(L_i) = \theta H(L_{i+1}) \sim \theta[H(L_i) + (L_{i+1} - L_i)H'(L_i)] \\ &\Leftrightarrow \quad L_{i+1} - L_i \sim e^{-L_i^2/2} H(L_i)(\frac{1}{\theta} - 1) \\ &\Leftrightarrow \quad L_{i+1} - L_i \sim \frac{1}{L_i}(\frac{1}{\theta} - 1) \ \text{ since } \ H(L_i) \underset{x\to\infty}{\sim} \frac{e^{x^2/2}}{x}\end{aligned}$$

If for example $\theta = 1/2$, the previous calculation suggests to take

$$L_{i+1} = L_i + \frac{1}{L_i}$$

and we deduce an equivalent of L_n for large values of n:

$$L_n \underset{n\to\infty}{\sim} \sqrt{2n}$$

from the following lemma

Lemma 5.1. *Let $(u_n)_n$ a real-valued sequence such that*

$$u_0 > 0 \quad \textit{and} \quad \forall n \in \mathbb{N}, \quad u_{n+1} = u_n + \frac{1}{u_n^\alpha}$$

with $\alpha > -1$. Then $\quad u_n \underset{n\to\infty}{\sim} [n(1+\alpha)]^{1/(\alpha+1)}$.

Proof. We only give a sketch of the proof of this classical analytical result. First, (u_n) diverges which comes from monotonicity and a fixed point argument. Then we look for a constant $\beta > 0$ such that $u_{n+1}^\beta - u_n^\beta$ converges. Finally, the result comes easily.

Cost of transition Now we deduce the asymptotic behavior of the cost of transition i.e. the cost for a particle starting from x to reach the next level $(x + \frac{1}{x})$ or 0:

$$\mathbb{E}_x(T_{0,x+\frac{1}{x}}) \underset{x\to+\infty}{\sim} (1 - e^{-1}) \log(x)$$

Proof. We have

$$\mathbb{E}_x(T_{0,x+\frac{1}{x}}) = -\int_0^x \frac{R(t)}{\rho(t)} dt + \frac{\int_0^{x+\frac{1}{x}} \frac{R(t)}{\rho(t)} dt}{\int_0^{x+\frac{1}{x}} \frac{1}{\rho(t)} dt} \int_0^x \frac{1}{\rho(t)} dt \tag{18}$$

Write $R(t) = R(\infty) - D(t)$ with $D(t) = \int_t^\infty \rho(u)du$; multiplying each side of the equation by $\int_0^{x+\frac{1}{x}} \frac{dt}{\rho(t)}$, we get

$$\begin{aligned}
\mathbb{E}_x(T_{0,x+\frac{1}{x}}) \int_0^{x+\frac{1}{x}} \frac{dt}{\rho(t)} &= \int_0^x \frac{dt}{\rho(t)} \int_0^{x+\frac{1}{x}} \frac{R(t)}{\rho(t)} dt - \int_0^{x+\frac{1}{x}} \frac{dt}{\rho(t)} \int_0^x \frac{D(t)}{\rho(t)} dt \\
&= -\int_0^x \frac{dt}{\rho(t)} \int_x^{x+\frac{1}{x}} \frac{D(t)}{\rho(t)} dt + \int_x^{x+\frac{1}{x}} \frac{dt}{\rho(t)} \int_0^x \frac{D(t)}{\rho(t)} dt
\end{aligned}$$

Using classical estimates, we are lead to the required result.

Remark 5.1. That result justifies our hypothesis that the costs of transition are asymptotically constant.

This type of results can also be deduced for the approximate Ornstein-Uhlenbeck process; but the analysis is quite more complicated.

5.2. Numerical application

We suppose the unknown drift is constant equal to μ_i on disjoint intervals and on each interval i

$$\mu_i \in [\mu_i^1, \mu_i^2].$$

We determine arbitrarily a "mean" drift constant on disjoint intervals taking on each interval i, $\mu_i^0 = (\mu_i^1 + \mu_i^2)/2$. We then determine the thresholds L_i such that the transition probabilities are equal to p and the optimal sampling numbers for that mean drift . Now we sample ρ_N processes according to

$$dX_t = -\mu(X_t)X_t dt + \sigma dW_t, \tag{19}$$

that we discretize in

$$X_{n+h} = -(\mu h - 1)X_n + \sigma G,$$

where h is the discretization step and G have the $\mathcal{N}(0, h)$ distribution. The sampling numbers used during that learning phase are the optimal sampling numbers for the mean drift previously determined. We get in that way estimates of the transition probabilities that allows us to estimate the optimal sampling numbers. We then sample $N - \rho_N$ processes according to the discretization of (19) with the estimated sampling numbers.

Remark 5.2. Simulation procedure
Due to the discretization of the process and since the drift is not constant, a bias is introduced at each change of regime of the drift. We propose here a procedure to reduce that bias. Let $(m_i)_i$ the levels of drift changes, $(X_n)_n$ the studied process. Suppose X_n such that $m_{i-1} < X_n < m_i$, then X_n evolve according to (19) with drift μ_i and

$$X_{n+1} = X_n - \mu_i h X_n + \sigma\sqrt{h}Z$$

where Z is a realization of the $\mathcal{N}(0,1)$ distribution.
If $X_{n+1} < m_i$, there is no problem: the process evolves from X_n according to (19) with drift μ_i and the next point X_{n+1} is still in the same regime of drift. And the simulation is simply continued. On the other hand, if $X_{n+1} > m_i$, we should have changed the drift between X_n and X_{n+1} which introduces the bias mentioned at the beginning of the remark. Thus we solve the second degree equation in X_{n+1}

$$X_{n+1} = X_n - \frac{\mu_i(m_i - X_n) + \mu_{i+1}(X_{n+1} - m_i)}{X_{n+1} - X_n} h X_n + \sigma\sqrt{h}Z$$

Finally, X_{n+1} is the obtained solution.

Application We aim at estimate the probability that the process governed by (19) (with $\sigma = 0.3$) and starting from $x = 0.1$ reaches level $L = 4$ before going to $L_0 = 0$. The values of the different parameters are reported in the following tabular

$[m_i, m_{i+1}]$	μ_i	μ_i^1	μ_i^2	μ_i^0	$[m_i, m_{i+1}]$	μ_i	μ_i^1	μ_i^2	μ_i^0
$[0, 1/2]$	0.05	0.05	0.06	0.055	$[1/2, 1]$	0.06	0.06	0.07	0.065
$[1, 3/2]$	0.07	0.07	0.08	0.075	$[3/2, 2]$	0.09	0.08	0.09	0.085
$[2, 5/2]$	0.09	0.09	0.10	0.095	$[5/2, 3]$	0.11	0.10	0.11	0.105
$[3, 7/2]$	0.11	0.11	0.12	0.115	$[7/2, 4]$	0.12	0.12	0.13	0.125

which corresponds to the target probability

$$\mathbb{P}_x(X_{T_{L_0,L}} = L) = 1.67679e - 08$$

The thresholds for that mean drift and $p = 0.2$ are given by

L_1	0.4778339279	L_2	1.406139441	L_3	2.073610760	L_4	2.480952409
L_5	2.785321189	L_6	3.039774162	L_7	3.257152634	L_8	3.453693339
L_9	3.632850394	L_{10}	3.795331329	L_{11}	3.949233228	L_{12}	4

The following table presents the results obtained with different methods, N ($= 10^4$, 10^5 or 10^6) processes generated at the beginning and 50 iterations. The abbreviations TSBSM, TSBSMimp, OSBSMimp, SALO and AACG mean respectively Two Steps Branching Splitting Model, Two Steps Branching Splitting Model improved with Remark 5.2, One Step Branching Splitting Model improved with Remark 5.2, Sequential Algorithm introduced by F.LeGland and N.Oudjane in [**?**] and Adaptive Algorithm introduced by F.Cérou and A.Guyader in [?].

	Nb of part.	Estimation	Error	Lgth CI 95%	Simul. cost
TSBSMimp	10^4	***	***	***	***
TSBSMimp	10^5	1.69e-08	1.21e-10	2.12e-09	1.19e+09
TSBSMimp	10^6	1.68e-08	1.21e-12	1.21e-09	1.19e+10
TSBSM	10^5	1.75e-08	7.68e-10	1.72e-09	1.21e+09
OSBSMimp	10^5	1.66e-08	1.32e-10	4.30e-09	1.30e+10
SALO	$H = 10^2$	1.89e-08	2.16e-09	4.08e-09	9.02e+07
SALO	$H = 10^3$	1.75e-08	7.61e-10	1.72e-09	9.18e+08
SALO	$H = 10^4$	1.67e-08	1.27e-10	4.07e-10	6.45e+09
AACG $p = 1/4$	10^5	1.05e-08	6.23e-09	2.06e-08	9.30e+08
AACG $p = 1/3$	10^5	1.14e-08	5.41e-09	1.66e-08	1.60e+09
AACG $p = 1/2$	10^5	1.22e-08	4.61e-09	2.56e-08	1.90e+09
AACG $p = 2/3$	10^5	1.98e-08	3.04e-09	2.97e-08	2.20e+09
AACG $p = 3/4$	10^5	1.86e-08	1.86e-09	2.86e-08	2.30e+09

In this technique, we run trajectories between two successive thresholds (which are given at the beginning of the simulation) until some number of successes H is reached.

In this technique, we run N trajectories until 0 is reached. The thresholds are then defined during the simulation in such a way that exactly a proportion p of these trajectories have succeeded to reach the level. To have a constant population, the trajectories that have failed are resampled on the successful ones.

6. Conclusion

We proposed in this paper a two-step algorithm based on the branching splitting model. A precise analysis shows that we shall dedicate asymptotically $\mu_s C^{2/3}$ particles to the learning phase and $C/C_{opt} - \mu_s C^{2/3}$ to the second phase, where C_{opt} is a constant and μ_s is derived by the optimization of the algorithm; i.e. assuming that the number of particles generated during the learning phase behaved like $\mu C^{1-\alpha}$, we should take $\alpha = 1/3$.

This result directly comes from the particular choice of the splitting numbers R_i's, solutions of an optimization problem. In whatever other algorithm i.e. if the R_i's are not taken in such a way, the optimal α would simply be equal to $1/2$, as one can expect to a priori. This emphasizes on the importance of choosing the transition probabilities and the splitting numbers as close as possible to the optimals and to have good estimates of the parameters chosen a priori. It also insists on the interest of having an adaptive algorithm in multiple phases that would be more efficient: evaluating at each step the parameters and continuing simulation according to these estimates. But, in such an algorithm, the calculation using martingale techniques rapidly become even more complex than those of this paper (see e.g. Appendix A and B). Moreover, to show that the adaptive algorithm works better than the one presented here would really be difficult and the gain resulting of such an approach really hard to quantify.

In terms of efficient algorithm, the best thing to do would be to estimate and to increment not only the splitting numbers but also the transition probabilities and then to move, before each phase, the thresholds according to these evaluations. But the precise analysis become more and more complex. Moreover such an algorithm lies on the hypothesis that one can move practically the thresholds which is in general not the case in practical settings.

Finally, it would also be worth to analyse non markovian models, multi-dimensional models and not only uni-dimensional models. In the formers, the probability to reach some level L_{i+1} starting from level L_i depends on the entry point in L_i. Such studies would be interesting and worth to analyse but once again really hard to lead.

Appendix A. Proof of Lemma 3.1

In this section, we want to prove Lemma 3.1. The first part of the lemma is clearly well-known and directly comes from the Chernoff's bounding method and optimization. Then, first of all,

$$\mathbb{E}\left(f(\widetilde{P})\right) = \mathbb{E}\left(f(\widehat{P})\mathbf{1}_{\widehat{P}\in[a,b]}\right) + \mathbb{E}\left(f(\widetilde{P})\mathbf{1}_{\widehat{P}\in[a,b]^c}\right)$$

But, since $\widetilde{P} \in [a,b]$ and f is bounded on $[a,b]$ by some constant M_f,

$$\mathbb{E}\left(f(\widetilde{P})\mathbf{1}_{\widehat{P}\in[a,b]^c}\right) \leq M_f\mathbb{P}\left(\widehat{P}\notin[a,b]\right) \leq 2M_f\exp\{-Nh(a,b)\}$$

by the first part of Lemma 3.1 and with $h(a,b) = \min\{l(a), l(b)\}$.

Then, one can notice that

$$\mathbb{E}\left(\widehat{P}\right) = P, \quad \mathrm{var}\left(\widehat{P}\right) = \frac{P(1-P)}{N}, \quad \mathbb{E}\left(\left(\widehat{P}-P\right)^3\right) = \frac{P(1-P)}{N^2}(1-2P)$$
$$\text{and } \mathbb{E}\left(\left(\widehat{P}-P\right)^4\right) = \frac{P(1-P)}{N^2}\left[(P^3+(1-P)^3)\frac{1}{N} + (1-\frac{1}{N})P(1-P)\right].$$

and since f is $\mathcal{C}^2$ on $[a,b]$, by a Taylor expansion,

$$f(\widehat{P})\mathbf{1}_{\widehat{P}\in[a,b]} = \left[f(P) + \left(\widehat{P}-P\right)f'(P) + \frac{\left(\widehat{P}-P\right)^2}{2}f''(P) + o\left(\left(\widehat{P}-P\right)^2\right)\right]\mathbf{1}_{\widehat{P}\in[a,b]}$$

and so

$$\mathbb{E}\left(f(\widehat{P})\mathbf{1}_{\widehat{P}\in[a,b]}\right) = \mathbb{E}\left(f(P) + \left(\widehat{P}-P\right)f'(P) + \frac{\left(\widehat{P}-P\right)^2}{2}f''(P) + o\left(\left(\widehat{P}-P\right)^2\right)\right)$$
$$-\mathbb{E}\left(\left[f(P) + \left(\widehat{P}-P\right)f'(P) + \frac{\left(\widehat{P}-P\right)^2}{2}f''(P) + o\left(\left(\widehat{P}-P\right)^2\right)\right]\mathbf{1}_{\widehat{P}\notin[a,b]}\right)$$

The first term in the right hand side is simply

$$f(P)+\mathbb{E}\left(\widehat{P}-P\right)f'(P)+\mathrm{var}\left(\widehat{P}\right)\frac{f''(P)}{2}+\mathbb{E}\left(o\left(\widehat{P}-P\right)^2\right) = f(P)+\frac{P(1-P)}{2N}f''(P)+o(\frac{1}{N})$$

Since $P \in [a,b]$ and f and its two first derivatives are bounded on $[a,b]$, the second term in the right hand side is bounded by some constant times $\mathbb{P}\left(\widehat{P} \notin [a,b]\right)$, itself bounded by $\exp\{-Nh(a,b)\}$, allowing us to conclude.

Appendix B. Cost asymptotic expression

In this appendix, we prove Theorem 4.1 giving the asymptotic expression of the cost:

$$C = N\left[C_{opt} + \frac{B_1}{\lambda N^{1-\alpha}} + B_2\frac{\lambda}{N^{\alpha}} + o\left(\frac{1}{N^{\alpha}}\right) + o\left(\frac{1}{N^{1-\alpha}}\right)\right]$$

with B_1, B_2 and C_{opt} constants given in Theorem 4.1.

Since C is given by (11), one needs to evaluate $\mathbb{E}(\widetilde{p}_i^{(1)})$ (which is obviously p_i) and $\mathbb{E}\left(\widetilde{r}_i\widetilde{p}_i^{(2)}\right)$. Now define $\varphi(x) = \sqrt{\frac{1}{x} - 1}$ and $\psi(x) = \sqrt{x(1-x)}$, then $\widetilde{R}_i = \frac{\varphi(\widetilde{P}_{i+1}^{(1)})}{\psi(\widetilde{P}_i^{(1)})}$. To derive the formula, we proceed by induction, successive conditionings and uses of Lemma 3.1. For $k = i$ (resp. $k = i-1\dots 2$), we

- condition by $\mathcal{F}_{2k}$ to isolate what happens from L_k and then apply Lemma 3.1 near P_{k+1} to the function ϕ (resp. to $\frac{\varphi}{\psi}$) and

$$\widetilde{P}_{k+1}^{(1)} = P_{k+1} + \frac{1}{\lambda N^{1-\alpha} r_k^0 \widetilde{p}_k^{(1)}} \sum_{j=1}^{\lambda N^{1-\alpha} r_k^0 \widetilde{p}_k^{(1)}} Ber'_j$$

  with $Ber'_j \underset{\mathcal{L}}{=} \mathrm{Ber}\left(P_{k+1}\right) - P_{k+1}$.
- condition by $\mathcal{F}_{2k-1}$ to isolate what happens from L_{k-1} after the first phase and then apply Lemma 3.1 near P_k to the function Id and

$$\widetilde{P}_k^{(2)} = P_k + \frac{1}{(N - \lambda N^{1-\alpha})\tilde{r}_k \widetilde{p}_{k-1}^{(2)}} \sum_{j=1}^{(N-\lambda N^{1-\alpha})\tilde{r}_k \widetilde{p}_{k-1}^{(2)}} Ber'_j$$

  with $Ber'_j \underset{\mathcal{L}}{=} \mathrm{Ber}\left(P_k\right) - P_k$.

At step $k = i$, we get

$$\begin{aligned}\mathbb{E}\left(\widetilde{r}_i\widetilde{p}_i^{(2)}\right) &= \mathbb{E}\left(\frac{\widetilde{r}_{i-1}}{\psi(\widetilde{p}_i^{(1)})}\widetilde{p}_i^{(2)}\mathbb{E}_{2i}\left(\varphi(\widetilde{P}_{i+1}^{(1)})\right)\right) = \varphi(P_{i+1}) \overbrace{\mathbb{E}\left(\frac{\widetilde{r}_{i-1}}{\psi(\widetilde{P}_i^{(1)})}\widetilde{p}_i^{(2)}\right)}^{=\mathbb{E}\left(\frac{\widetilde{r}_{i-1}}{\psi(\widetilde{P}_i^{(1)})}\widetilde{p}_{i-1}^{(2)}\mathbb{E}_{2i-1}\left(\widetilde{P}_i^{(2)}\right)\right)=P_i\mathbb{E}\left(\frac{\widetilde{r}_{i-1}}{\psi(\widetilde{P}_i^{(1)})}\widetilde{p}_{i-1}^{(2)}\right)+o\left(\frac{1}{N^{1-\alpha}}\right)} \\ &+ \frac{P_{i+1}(1-P_{i+1})}{2\lambda N^{1-\alpha}r_i^0}\varphi''(P_{i+1})\underbrace{\mathbb{E}\left(\frac{\widetilde{r}_{i-1}}{\psi(\widetilde{P}_i^{(1)})}\frac{\widetilde{p}_i^{(2)}}{\widetilde{p}_i^{(1)}}\right)}_{=\frac{r_{i-1}}{\psi(P_i)}+o\left(\frac{1}{N^{1-\alpha}}\right)} + o\left(\frac{1}{N^{1-\alpha}}\right)\end{aligned}$$

and then proceeding for k from $i-1$ to 2 leads to

$$\begin{aligned}\mathbb{E}\left(\widetilde{r}_i\widetilde{p}_i^{(2)}\right) &= \frac{r_i}{R_1}\frac{p_i}{P_1}\varphi(P_2)\mathbb{E}\left(\frac{\widetilde{P}_1^{(2)}}{\psi(\widetilde{P}_1^{(1)})}\right) + \frac{r_i p_i}{2\lambda N^{1-\alpha}}\sum_{k=1}^{i-1}\frac{P_{k+1}(1-P_{k+1})}{r_k^0 R_k R_{k+1}}\frac{\varphi(P_{k+2})}{p_k\psi(P_k)}\left(\frac{\varphi}{\psi}\right)''(P_{k+1}) \\ &+ \frac{P_{i+1}(1-P_{i+1})}{2\lambda N^{1-\alpha}r_i^0}\varphi''(P_{i+1})\frac{r_{i-1}}{\psi(P_i)} + o\left(\frac{1}{N^{1-\alpha}}\right)\end{aligned}$$

To derive the required expression, it remains to apply Lemma 3.1 to $(x,y) \mapsto \frac{y}{\psi(x)}$.

Appendix C. Variance asymptotic expression

In this appendix, we prove Theorem 4.2 giving the asymptotic expression of the variance:

$$var(\widetilde{P}) = \frac{\mathbb{P}(L)^2}{N}\left[V_{opt} + \frac{1}{\lambda N^{1-\alpha}}A_1 + \frac{\lambda}{N^\alpha}A_2 + \frac{\lambda^2}{N^{2\alpha}}A_3 + o\left(\frac{1}{N^{1-\alpha}}\right) + o\left(\frac{1}{N^{2\alpha}}\right)\right]$$

with A_1, A_2, A_3 and C_{opt} constants given in Theorem 4.2.

To prove the result, we establish a recursive relation between $\text{var}\left(\widetilde{P}\right) = \text{var}(\widetilde{p}_{M+1})$ and $\text{var}(\widetilde{p}_M)$ and we conclude by iterating this relation. During the proof, in order to lighten drafting, we use the symbol $\approx$ for any asymptotic expansion at the order $1/N^{(1+\alpha)\wedge(2-\alpha)}$. Now for $k = 1..M$, let Z_k the ratio between the numbers of particles having reached level L_k in phases 1 and 2:

$$Z_k = \frac{\widetilde{p}_k^{(1)}r_k^0}{\widetilde{p}_k^{(2)}\widetilde{r}_k} =: \frac{\overline{Z}_k}{\widetilde{R}_k} = \overline{Z}_k\frac{\psi(\widetilde{P}_k^{(1)})}{\varphi(\widetilde{P}_{k+1}^{(1)})} \qquad \text{then} \qquad \widetilde{P}_k = \widetilde{P}_k^{(2)}\frac{1 + \frac{\lambda}{N^\alpha}\left(Z_{k-1}\frac{\widetilde{P}_k^{(1)}}{\widetilde{P}_k^{(2)}} - 1\right)}{1 + \frac{\lambda}{N^\alpha}(Z_{k-1} - 1)}$$

We proceed as in Appendix A by successive conditionings and several applications of Lemma 3.1. Since the estimate $\widetilde{P} = \widetilde{p}_{M+1}$ of $\mathbb{P}(L)$ is expressed as a product

$\widetilde{P}_1 \ldots \widetilde{P}_{M+1}$, conditionings appears to be particularly efficient to deduce the required recursive relation. A first conditioning by $\mathcal{F}_{2M+1}$ leads to

$$\begin{aligned}
&\operatorname{var}\left(\widetilde{P}\right) = \operatorname{var}(\widetilde{p}_{M+1}) \\
&= \operatorname{var}\left(\frac{\widetilde{p}_M}{\widetilde{p}_M^{(2)}\widetilde{r}_M\left\{1+\frac{\lambda}{N^\alpha}\left(Z_M-1\right)\right\}}\left\{\frac{\lambda}{N^\alpha}\widetilde{p}_{M+1}^{(1)}r_M^0 + \left(1-\frac{\lambda}{N^\alpha}\right)\widetilde{p}_M^{(2)}\widetilde{r}_M\mathbb{E}_{2M+1}(\widetilde{P}_{M+1}^{(2)})\right\}\right) \\
&+\mathbb{E}\left(\left[\frac{\widetilde{p}_M}{\widetilde{p}_M^{(2)}\widetilde{r}_M\left\{1+\frac{\lambda}{N^\alpha}\left(Z_M-1\right)\right\}}\right]^2\left\{\left(1-\frac{\lambda}{N^\alpha}\right)\widetilde{p}_M^{(2)}\widetilde{r}_M\right\}^2\operatorname{var}_{2M+1}\left(\widetilde{P}_{M+1}^{(2)}\right)\right) \\
&\approx P_{M+1}^2\operatorname{var}\left(\widetilde{p}_M\left\{1+\frac{\lambda}{N^\alpha}Z_M\left(\frac{\widetilde{P}_{M+1}^{(1)}}{\widetilde{P}_{M+1}^{(1)}}-1\right)\left[1-\frac{\lambda}{N^\alpha}\left(Z_M-1\right)\right]\right\}\right) \\
&+\frac{P_{M+1}(1-P_{M+1})}{N}\left(1-\frac{\lambda}{N^\alpha}\right)\mathbb{E}\left(\frac{\widetilde{p}_M^2}{\widetilde{p}_M^{(2)}\widetilde{r}_M}\left\{1-2\frac{\lambda}{N^\alpha}\left(Z_M-1\right)+3\frac{\lambda^2}{N^{2\alpha}}\left(Z_M-1\right)^2\right\}\right)
\end{aligned}$$

using $\mathbb{E}_{2M+1}(\widetilde{P}_{M+1}^{(2)}) = P_{M+1} + o\left(\frac{1}{N}\right)$ and $\operatorname{var}_{2M+1}\left(\widetilde{P}_{M+1}^{(2)}\right) = \frac{P_{M+1}(1-P_{M+1})}{N\left(1-\frac{\lambda}{N^\alpha}\right)\widetilde{p}_M^{(2)}\widetilde{r}_M} + o\left(\frac{1}{N}\right)$ (by Lemma 3.1) and $(1+x)^\alpha \underset{\alpha\to 0}{\approx} (1+\alpha x+\alpha(\alpha+1)/2x^2)$;

$$\begin{aligned}
&\approx P_{M+1}^2\operatorname{var}\left(\widetilde{p}_M\left\{1+\frac{\lambda}{N^\alpha}\overline{Z}_M\psi(\widetilde{P}_M^{(1)})\mathbb{E}_{2M}\left(f(\widetilde{P}_{M+1}^{(1)})\right)\right\}\right) \\
&+P_{M+1}^2\mathbb{E}\left(\widetilde{p}_M^2\frac{\lambda^2}{N^{2\alpha}}\left[\overline{Z}_M\psi(\widetilde{P}_M^{(1)})\right]^2\operatorname{var}_{2M}\left(f(\widetilde{P}_{M+1}^{(1)})\right)\right) \\
&+\frac{P_{M+1}(1-P_{M+1})}{N}\left(1-\frac{\lambda}{N^\alpha}\right)\mathbb{E}\left(\frac{\widetilde{p}_M^2}{\widetilde{p}_M^{(2)}\widetilde{r}_{M-1}}\psi(\widetilde{P}_M^{(1)})\mathbb{E}_{2M}\left(g(\widetilde{P}_{M+1}^{(1)})\right)\right)
\end{aligned}$$

where the last step is the result of two conditionings by $\mathcal{F}_{2M}$ and with

$$f(x) = \frac{1}{\varphi(x)}\left(\frac{x}{P_{M+1}}-1\right)\left\{1-\frac{\lambda}{N^\alpha}\left(\overline{Z}_M\frac{\psi(\widetilde{P}_M^{(1)})}{\varphi(x)}-1\right)\right\}$$

and

$$g(x) = \frac{1}{\varphi(x)}\left\{1-2\frac{\lambda}{N^\alpha}\left(\overline{Z}_M\frac{\psi(\widetilde{P}_M^{(1)})}{\varphi(x)}-1\right)+3\frac{\lambda^2}{N^{2\alpha}}\left(\overline{Z}_M\frac{\psi(\widetilde{P}_M^{(1)})}{\varphi(x)}-1\right)^2\right\}$$

1. <u>Calculation of the first term denoted T_1</u>: Applying Lemma 3.1 to f, we get

$$
\begin{aligned}
T_1 &\approx P_{M+1}^2 \mathrm{var}\left(\widetilde{p}_M \left\{1+\frac{1}{N}\frac{\psi(\widetilde{P}_M^{(1)})\varphi(P_{M+1})^{-1}}{\widetilde{p}_M^{(2)}\widetilde{r}_{M-1}P_{M+1}}\left\{1-\frac{\lambda}{N^\alpha}\left(2\overline{Z}_M\frac{\psi(\widetilde{P}_M^{(1)})}{\varphi(P_{M+1})}-1\right)\right\}\right\}\right)\\
&\approx P_{M+1}^2 \mathrm{var}\left(\widetilde{p}_M\right)\\
&\quad+\frac{1}{4N^2\varphi(P_{M+1})^2}\mathrm{var}\left(\frac{\widetilde{p}_M\psi(\widetilde{P}_M^{(1)})}{\widetilde{p}_M^{(2)}\widetilde{r}_{M-1}}\left\{1-\frac{\lambda}{N^\alpha}\left(2\overline{Z}_M\frac{\psi(\widetilde{P}_M^{(1)})}{\varphi(P_{M+1})}-1\right)\right\}\right)\\
&\quad+\frac{P_{M+1}}{N\varphi(P_{M+1})}\mathrm{cov}\left(\widetilde{p}_M,\frac{\widetilde{p}_M\psi(\widetilde{P}_M^{(1)})}{\widetilde{p}_M^{(2)}\widetilde{r}_{M-1}}\left\{1-\frac{\lambda}{N^\alpha}\left(2\overline{Z}_M\frac{\psi(\widetilde{P}_M^{(1)})}{\varphi(P_{M+1})}-1\right)\right\}\right)
\end{aligned}
$$

The second term (which is in $\frac{1}{N^2}$) being negligible, we only study the term involving the covariance equals to

$$
\begin{aligned}
&\mathbb{E}\left(\widetilde{p}_M^2\frac{\psi(\widetilde{P}_M^{(1)})}{\widetilde{p}_M^{(2)}\widetilde{r}_{M-1}}\left\{1-\frac{\lambda}{N^\alpha}\left(2\overline{Z}_M\frac{\psi(\widetilde{P}_M^{(1)})}{\varphi(P_{M+1})}-1\right)\right\}\right)\\
&-\mathbb{E}\left(\widetilde{p}_M\right)\mathbb{E}\left(\widetilde{p}_M\frac{\psi(\widetilde{P}_M^{(1)})}{\widetilde{p}_M^{(2)}\widetilde{r}_{M-1}}\left\{1-\frac{\lambda}{N^\alpha}\left(2\overline{Z}_M\frac{\psi(\widetilde{P}_M^{(1)})}{\varphi(P_{M+1})}-1\right)\right\}\right)=:T_{11}-T_{12}T_{13}
\end{aligned}
$$

and we show that it is null. The integration procedure being the same for all the expectations introduced in this proof, we only detail the calculation of $\mathbb{E}\left(\widetilde{p}_M\right)$ in order to lighten drafting:

$$
\begin{aligned}
T_{12} &= \mathbb{E}\left(\widetilde{p}_M\right)=\mathbb{E}\left(\widetilde{p}_{M-1}\widetilde{P}_M^{(2)}\left\{\frac{1+\frac{\lambda}{N^\alpha}\left(Z_{M-1}\frac{\widetilde{P}_M^{(1)}}{\widetilde{P}_M^{(2)}}-1\right)}{1+\frac{\lambda}{N^\alpha}\left(Z_{M-1}-1\right)}\right\}\right)\\
&=\mathbb{E}\left(\widetilde{p}_{M-1}\left\{1+\frac{\lambda}{N^\alpha}\left(Z_{M-1}-1\right)\right\}^{-1}\mathbb{E}_{2M-1}\left(u_M(\widetilde{P}_M^{(2)})\right)\right)
\end{aligned}
$$

conditioning by $\mathcal{F}_{2M-1}$ and noting $u_M(x)=x\left\{1+\frac{\lambda}{N^\alpha}\left(Z_{M-1}\frac{\widetilde{P}_M^{(1)}}{x}-1\right)\right\}$. But by Lemma 3.1 (the second term being negligible),

$$
\begin{aligned}
\mathbb{E}_{2M-1}\left(u_M(\widetilde{P}_M^{(2)})\right) &= u_M(P_M)+\frac{P_M(1-P_M)}{2N\left(1-\frac{\lambda}{N^\alpha}\right)}u_M''(P_M)+o\left(\frac{1}{N}\right)\\
&= P_M\left\{1+\frac{\lambda}{N^\alpha}\left(Z_{M-1}\frac{\widetilde{P}_M^{(1)}}{P_M}-1\right)\right\}+o\left(\frac{1}{N^\alpha}\right)
\end{aligned}
$$

Let now

$$v_M(x) = \left\{1 - \frac{\lambda}{N^\alpha}\left(\overline{Z}_{M-1}\frac{\psi(\widetilde{P}_{M-1}^{(1)})}{\varphi(x)} - 1\right) + \frac{\lambda^2}{2N^{2\alpha}}\left(\overline{Z}_{M-1}\frac{\psi(\widetilde{P}_{M-1}^{(1)})}{\varphi(x)} - 1\right)^2\right\}$$
$$\left\{1 + \frac{\lambda}{N^\alpha}\left(\overline{Z}_{M-1}\frac{\psi(\widetilde{P}_{M-1}^{(1)})}{\varphi(x)}\frac{x}{P_M} - 1\right)\right\}$$

and apply Lemma 3.1 to v_M to get

$$\mathbb{E}_{2(M-1)}\left(v_M(\widetilde{P}_M^{(1)})\right) = v_M(P_M) + \frac{P_M(1-P_M)}{2\lambda N^{1-\alpha}}v_M''(P_M) + o\left(\frac{1}{N^{1-\alpha}}\right) = 1 + o\left(\frac{1}{N}\right).$$

Then by a conditioning by $\mathcal{F}_{2(M-1)}$, one can deduce

$$T_{12} \approx P_M\mathbb{E}\left(\widetilde{p}_{M-1}\mathbb{E}_{2(M-1)}\left(v_M(\widetilde{P}_M^{(1)})\right)\right) \approx P_M\mathbb{E}\left(\widetilde{p}_{M-1}\right) \approx \cdots \approx p_M.$$

Making successive conditionings and applications of Lemma 3.1, T_{11} becomes

$$\begin{aligned}
T_{11} &\approx \frac{P_M\cdots P_2}{R_{M-1}\cdots R_2}\frac{\psi(P_M)}{\varphi(P_2)}\mathbb{E}\left(\widetilde{P}_1^2\frac{\psi(\widetilde{P}_1^{(1)})}{\widetilde{P}_1^{(2)}}\left\{1 - \frac{\lambda}{N^\alpha}\left(2\frac{\widetilde{P}_1^{(1)}r_M^0}{\widetilde{P}_1^{(2)}}\frac{\psi(\widetilde{P}_1^{(1)})}{\varphi(P_2)} - 1\right)\right\}\right)\\
&= \frac{p_M\psi(P_M)}{r_{M-1}}\left[\left\{1 - \frac{\lambda}{N^\alpha}\left(2\frac{r_M^0}{r_M} - 1\right)\right\} - \frac{1/8}{\lambda N^{1-\alpha}}\frac{1}{P_1(1-P_1)}\right] + o\left(\frac{1}{N^{1-\alpha}}\right)
\end{aligned}$$

by an application of Lemma 3.1 to Φ, defined by

$$\Phi(x,y) = y\psi(x)\left\{1 + 2\frac{\lambda}{N^\alpha}\left(\frac{x}{y} - 1\right) + \frac{\lambda^2}{N^{2\alpha}}\left(\frac{x}{y} - 1\right)^2\right\}\left\{1 - \frac{\lambda}{N^\alpha}\left(\frac{2r_M^0}{\varphi(P_2)}\frac{x\psi(x)}{y} - 1\right)\right\},$$

to obtain

$$\begin{aligned}
\mathbb{E}\left(\Phi\left(\widetilde{P}_1^{(1)}, \widetilde{P}_1^{(2)}\right)\right) &= \Phi(P_1,P_1) + P_1(1-P_1)\left\{\frac{\Phi_x''(P_1,P_1)}{2\lambda N^{1-\alpha}} + \frac{\Phi_y''(P_1,P_1)}{2N(1-\lambda/N^\alpha)}\right\} + o\left(\frac{1}{N^{1-\alpha}}\right)\\
&= P_1\psi(P_1)\left\{1 - \frac{\lambda}{N^\alpha}\left(2\frac{r_M^0}{r_M} - 1\right)\right\} + P_1(1-P_1)\frac{\Phi_x''(P_1,P_1)}{2\lambda N^{1-\alpha}} + o\left(\frac{1}{N^{1-\alpha}}\right)
\end{aligned}$$

In the same way, we get

$$T_{13} = \frac{\psi(P_M)}{r_{M-1}}\left[\left\{1 - \frac{\lambda}{N^\alpha}\left(2\frac{r_M^0}{r_M} - 1\right)\right\} - \frac{1/8}{\lambda N^{1-\alpha}}\frac{1}{P_1(1-P_1)}\right] + o\left(\frac{1}{N^{1-\alpha}}\right)$$

from which we deduce the nullity of the covariance term. Finally, $T_1 \approx P_{M+1}^2\mathrm{var}\left(\widetilde{p}_M\right)$.

2. Calculation of the second term denoted T_2: Making successive conditionings and

applications of Lemma 3.1, it comes

$$\begin{aligned} T_2 &\approx \frac{\lambda}{N^{1+\alpha}}(1-P_{M+1})P_{M+1}\cdots P_2 \frac{r_M^0}{(R_M\cdots R_2)^2\varphi(P_2)^2} \\ &\quad \mathbb{E}\left(\widetilde{P}_1^2 \frac{\widetilde{P}_1^{(1)}\psi(\widetilde{P}_1^{(1)})^2}{\widetilde{P}_1'^2}\left\{1-2\frac{\lambda}{N^\alpha}\left(\frac{r_M^0}{R_M\cdots R_2}\frac{\psi(\widetilde{P}_1^{(1)})}{\varphi(P_2)}\frac{\widetilde{P}_1^{(1)}}{\widetilde{P}_1^{(2)}}-1\right)\right\}\right) \\ &= \frac{\lambda}{N^{1+\alpha}}(1-P_{M+1})p_{M+1}\frac{r_M^0}{r_M^2}\left\{1-2\frac{\lambda}{N^\alpha}\left(\frac{r_M^0}{r_M}-1\right)\right\}+o\left(\frac{1}{N^{1+\alpha}}\right) \end{aligned}$$

where the last step comes from an application of Lemma 3.1 to Ψ, defined by

$$\Psi(x,y) = x\psi(x)^2\left\{1+\frac{\lambda}{N^\alpha}\left(\frac{x}{y}-1\right)\right\}\left\{1-2\frac{\lambda}{N^\alpha}\left(\frac{r_M^0}{R_M\cdots R_2}\frac{\psi(x)}{\varphi(P_2)}\frac{x}{y}-1\right)\right\},$$

just keeping the first term.

3. <u>Calculation of the third term denoted T_3</u>: Applying Lemma 3.1 to g,

$$\begin{aligned} T_3 \approx\ & \frac{1}{2\lambda N^{2-\alpha}}\frac{P_{M+1}-1/4}{r_M r_M^0}+\frac{P_{M+1}(1-P_{M+1})}{N}\left(1-\frac{\lambda}{N^\alpha}\right)\mathbb{E}\left(\frac{\widetilde{p}_M^2}{\widetilde{p}_M'\widetilde{r}_{M-1}}\frac{\psi(\widetilde{P}_M^{(1)})}{\varphi(P_{M+1})}\right. \\ & \left.\left\{1-2\frac{\lambda}{N^\alpha}\left(\frac{r_M^0}{R_M\cdots R_2}\frac{\psi(\widetilde{P}_1^{(1)})}{\varphi(P_2)}\frac{\widetilde{P}_1^{(1)}}{\widetilde{P}_1^{(2)}}-1\right)+3\frac{\lambda^2}{N^{2\alpha}}\left(\frac{r_M^0}{R_M\cdots R_2}\frac{\psi(\widetilde{P}_1^{(1)})}{\varphi(P_2)}\frac{\widetilde{P}_1^{(1)}}{\widetilde{P}_1^{(2)}}-1\right)^2\right\}\right) \end{aligned}$$

But the second term of the right hand side becomes after successive conditionings and applications of Lemma 3.1,

$$\begin{aligned} & \frac{1-P_{M+1}}{N}\left(1-\frac{\lambda}{N^\alpha}\right)\frac{P_{M+1}\cdots P_2}{R_M\cdots R_2\varphi(P_2)}\mathbb{E}\left(\widetilde{P}_1^2\frac{\psi(\widetilde{P}_1^{(1)})}{\widetilde{P}_1'}\left\{1-2\frac{\lambda}{N^\alpha}\left(\frac{r_M^0}{R_M\cdots R_2}\frac{\psi(\widetilde{P}_1^{(1)})}{\varphi(P_2)}\frac{\widetilde{P}_1^{(1)}}{\widetilde{P}_1^{(2)}}-1\right)\right.\right. \\ & \qquad\qquad \left.\left.+3\frac{\lambda^2}{N^{2\alpha}}\left(\frac{r_M^0}{R_M\cdots R_2}\frac{\psi(\widetilde{P}_1^{(1)})}{\varphi(P_2)}\frac{\widetilde{P}_1^{(1)}}{\widetilde{P}_1^{(2)}}-1\right)^2\right\}\right) \\ \approx\ & \frac{p_{M+1}(1-P_{M+1})}{Nr_M}\left(1-\frac{\lambda}{N^\alpha}\right)\left\{1-2\frac{\lambda}{N^\alpha}\left(\frac{r_M^0}{r_M}-1\right)+3\frac{\lambda^2}{N^{2\alpha}}\left(\frac{r_M^0}{r_M}-1\right)^2\right\}-\frac{1-P_{M+1}}{8\lambda N^{2-\alpha}}\frac{p_{M+1}}{r_M P_1(1-P_1)} \end{aligned}$$

where the last step comes from an application of Lemma 3.1 to Γ, defined by

$$\begin{aligned} \Gamma(x,y) =\ & y\psi(x)\left\{1+2\frac{\lambda}{N^\alpha}\left(\frac{x}{y}-1\right)+\frac{\lambda^2}{N^{2\alpha}}\left(\frac{x}{y}-1\right)^2\right\} \\ & \left\{1-2\frac{\lambda}{N^\alpha}\left(\frac{r_M^0}{R_M\cdots R_2}\frac{\psi(x)}{\varphi(P_2)}\frac{x}{y}-1\right)+3\frac{\lambda^2}{N^{2\alpha}}\left(\frac{r_M^0}{R_M\cdots R_2}\frac{\psi(x)}{\varphi(P_2)}\frac{x}{y}-1\right)^2\right\} \end{aligned}$$

just keeping the two first terms.

From all these results, we deduce the recursive relation between $\mathrm{var}\left(\widetilde{p}_{M+1}\right)\left(=\mathrm{var}\left(\widetilde{P}\right)\right)$

and $\mathrm{var}(\widetilde{p}_M)$

$$\begin{aligned}\mathrm{var}\left(\widetilde{P}\right) &\approx P_{M+1}^2 \mathrm{var}(\widetilde{p}_M) + \frac{\lambda}{N^{1+\alpha}}(1-P_{M+1})p_{M+1}\frac{r_M^0}{r_M^2}\left\{1-2\frac{\lambda}{N^\alpha}\left(\frac{r_M^0}{r_M}-1\right)\right\} \\ &+\frac{p_{M+1}(1-P_{M+1})}{Nr_M}\left(1-\frac{\lambda}{N^\alpha}\right)\left\{1-2\frac{\lambda}{N^\alpha}\left(\frac{r_M^0}{r_M}-1\right)+3\frac{\lambda^2}{N^{2\alpha}}\left(\frac{r_M^0}{r_M}-1\right)^2\right\} \\ &+\frac{1}{2\lambda N^{2-\alpha}}\frac{1}{r_M}\left[\frac{P_{M+1}-1/4}{r_M^0}-\frac{p_{M+1}}{4}\frac{1-P_{M+1}}{P_1(1-P_1)}\right]\end{aligned}$$

It just remains now to evaluate $\mathrm{var}(\widetilde{p}_1)\left(=\mathrm{var}\left(\widetilde{P}_1\right)\right)$ to initialize the induction: applying Lemma 3.1,

$$\begin{aligned}\mathrm{var}\left(\widetilde{P}_1\right) &= \mathrm{var}\left(\widetilde{P}_1^{(2)}\left\{1+\frac{\lambda}{N^\alpha}\left(\frac{\widetilde{P}_1^{(1)}}{\widetilde{P}_1^{(2)}}-1\right)\right\}\right)+o\left(\frac{1}{N^{1-\alpha}}\right) \\ &= P_1(1-P_1)\left[\frac{\Delta_x'^2(P_1,P_1)}{\lambda N^{1-\alpha}}+\frac{\Delta_y'^2(P_1,P_1)}{N\left(1-\frac{\lambda}{N^\alpha}\right)}\right]+o\left(\frac{1}{N^{1-\alpha}}\right) \\ &= P_1(1-P_1)\left[\frac{1}{\lambda N^{1-\alpha}}\frac{\lambda^2}{N^{2\alpha}}+\frac{\left(1-\frac{\lambda}{N^\alpha}\right)^2}{N\left(1-\frac{\lambda}{N^\alpha}\right)}\right]+o\left(\frac{1}{N^{1-\alpha}}\right) \\ &= \frac{P_1(1-P_1)}{N}+o\left(\frac{1}{N^{1-\alpha}}\right)\end{aligned}$$

with $\Delta(x,y)=y\left(1+\frac{\lambda}{N^\alpha}\left(\frac{x}{y}-1\right)\right)$ and we conclude by an induction.

Appendix D. Proof of Lemma 4.1

In this section, we want to prove Lemma 4.1. We only detail the case $\alpha=\frac{1}{3}$ (the other results can easily be deduced by the same technique). First, since N is given by $G=G_f$ and by the asymptotic expression of G, N satisfies

$$N-N^{2/3}f_1(\lambda)+N^{1/3}f_4(\lambda)\sim G_f \tag{20}$$

and so we can consider that

$$N\sim G_f+\mu G_f^{2/3}+\nu G_f^{1/3}=G_f\left[1+\frac{\mu}{G_f^{1/3}}+\frac{\nu}{G_f^{2/3}}\right]$$

for some functions μ and ν to be determined. But injecting that value in (20), we finally get

$$G_f\sim G_f+[\mu-f_1(\lambda)]\,G_f^{2/3}+\left[\nu+f_4(\lambda)-\frac{2}{3}f_1(\lambda)\mu\right]G_f^{1/3}$$

from which we deduce that $\mu = f_1(\lambda)$ and $\nu = \frac{2}{3}f_1(\lambda)^2 - f_4(\lambda)$ and

$$N \sim G_f \left[1 + \frac{f_1(\lambda)}{G_f^{1/3}} + \frac{\frac{2}{3}f_1(\lambda)^2 - f_4(\lambda)}{G_f^{2/3}}\right].$$

Now injecting that N in the asymptotic expression of F in Lemma 4.1, we get

$$\begin{aligned}
F(\lambda, N) &\sim \frac{1}{G_f}\left[1 - \frac{\mu}{G_f^{1/3}} - \frac{\nu}{G_f^{2/3}} + \mu^2 \frac{1}{G_f^{2/3}}\right] + \frac{f_1(\lambda)}{G_f^{4/3}}\left[1 - \frac{4}{3}\frac{\mu}{G_f^{1/3}}\right] + \frac{(f_2+f_3)(\lambda)}{G_f^{5/3}} \\
&\sim \frac{1}{G_f} + \frac{1}{G_f^{4/3}}\left[f_1(\lambda) - \mu\right] + \frac{1}{G_f^{5/3}}\left[(f_2+f_3)(\lambda) - \frac{4}{3}\mu f_1(\lambda) - \nu + \mu^2\right] \\
&\sim \frac{1}{G_f} + \frac{1}{G_f^{5/3}}\left[(f_2+f_3)(\lambda) + f_4(\lambda) - f_1(\lambda)^2\right]
\end{aligned}$$

from which we deduce the required result for $\alpha = \frac{1}{3}$.

References

[1] David Aldous. *Probability approximations via the Poisson clumping heuristic*, volume 77 of *Applied Mathematical Sciences*. Springer-Verlag, New York, 1989.

[2] Davis Aldous and Umesh V. Vazirani. "go with the winners" algorithms. *IEEE Symposium on Fundations of Computer Science*, (7):492–501, 1994.

[3] Krishna B. Athreya and Peter E. Ney. *Branching processes*. Springer-Verlag, New York, 1972. Die Grundlehren der mathematischen Wissenschaften, Band 196.

[4] Frédéric Cérou and Arnaud Guyader. Adaptive multilevel splitting for rare event analysis. *Rapport de recherche de l'INRIA - Rennes , Equipe : ASPI*. 2005.

[5] Peter Tjerk de Boer. Analysis and efficient simulation of queueing models of telecommunication systems. 2000.

[6] Pierre Del Moral. *Feynman-Kac formulae*. Probability and its Applications (New York). Springer-Verlag, New York, 2004. Genealogical and interacting particle systems with applications.

[7] Persi Diaconis and Susan Holmes. Three examples of Monte-Carlo Markov chains: at the interface between statistical computing, computer science, and statistical

mechanics. In *Discrete probability and algorithms (Minneapolis, MN, 1993)*, volume 72 of *IMA Vol. Math. Appl.*, pages 43–56. Springer, New York, 1995.

[8] Arnaud Doucet, Nando de Freitas, and Neil Gordon. An introduction to sequential Monte Carlo methods. In *Sequential Monte Carlo methods in practice*, Stat. Eng. Inf. Sci., pages 3–14. Springer, New York, 2001.

[9] Theodore E. Harris. *The theory of branching processes.* Dover Phoenix Editions. Dover Publications Inc., Mineola, NY, 2002. Corrected reprint of the 1963 original [Springer, Berlin; MR **29** #664].

[10] Philip Heidelberger. Fast simulation of rare events in queueing and reliability models. *ACM Transactions on Modeling and Simulation*, 5(1):43–85, 1995.

[11] Mark Jerrum and Alistair Sinclair. *The Markov chain Monte Carlo method: an approach to approximate counting and integration.* Approximation Algorithms for NP-hard Problems. Dorit Hochbaum, 1997.

[12] Agnès Lagnoux. Rare event simulation. *Probab. Engrg. Inform. Sci.*, 20(1):45–66, 2006.

[13] Agnès Lagnoux-Renaudie. Effective branching splitting method under cost constraint, *Submitted to Stochastic Processes and their Applications.*

[14] François LeGland and Nadia Oudjane. *A Sequential Particle Algorithm that Keeps the Particle System Alive.* Stochastic Hybrid Systems : Theory and Safety Critical Applications. Springer-Verlag, Berlin, 2006. Lecture Notes in Control and Information Sciences 337.

[15] Russell Lyons. Probability on trees and networks. *Preprint.* http://mypage.iu.edu/ rdlyons/. 2002.

[16] John S. Sadowsky. On Monte Carlo estimation of large deviations probabilities. *Ann. Appl. Probab.*, 6(2):399–422, 1996.

[17] Manuel Villén-Altamirano and José Villén-Altamirano. Restart: a method for accelerating rare event simulations. pages 71–76. North-Holland, 1991.

[18] Manuel Villén-Altamirano and José Villén-Altamirano. Restart: An efficient and general method for fast simulation of rare event. *Technical report 7*. 1997.

Acknowledgements

I gratefully acknowledge my supervisors Dominique Bakry and Pascal Lezaud for their helpful suggestions and comments. I also thank the reviewer for a critical and attentive reading of the manuscript and precious detailed comments.