Effective branching splitting method under cost contraint

Agnès Lagnoux-Renaudie

Laboratoire de Statistique et Probabilités
Université Paul Sabatier
31062 Toulouse Cedex 4 France
Phone number : +33 (0)5 61 55 76 69
Fax : +33 (0)5 61 55 60 89

Résumé

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 during the simulation. Given the cost, the optimization of the algorithm suggests to sample a number of subtrials which may be non-integer and even unknown but estimated. To avoid this problem, we present in this paper three different approaches which provide precise estimates of the relative error between $\mathbb{P}(A)$ and its estimator.

Key words: splitting method, simulation, cost function, Laplace transform, Galton-Watson, branching processes, iterated functions, rare event
MSC : 65U05, 44A10, 60J80

1 Introduction

1.1 General settings

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

Email address: `lagnoux@cict.fr` (Agnès Lagnoux-Renaudie).
URL: `http ://www.lsp.ups-tlse.fr/Fp/Lagnoux` (Agnès Lagnoux-Renaudie).

Preprint submitted to Stochastic Processes and their Applications26 octobre 2007

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 considered and then in using mathematical (Aldous [1] and Sadowsky [26]) 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, the first one is based on importance sampling (IS), e.g. de Boer [7] and Heidelberger [20]. Fundamentally, IS is based on the notion of modifying the underlying probability distribution in such a way that the rare event occurs much more frequently. But the use of IS requires a deep knowledge of the studied system. An alternative way is to use trajectory splitting, based on the idea that 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. In contrast to IS type algorithms, the step-by-step evolution of the system follows the original probability measure. Thus we consider a decreasing sequence of events B_i leading to the rare event A :

$$A = B_{M+1} \subset B_M \subset \ldots \subset B_1 .$$

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

$$\mathbb{P}(A) = \mathbb{P}(A|B_M)\mathbb{P}(B_M|B_{M-1}) \ldots \mathbb{P}(B_2|B_1)\mathbb{P}(B_1), \tag{1}$$

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

The principle of the algorithm is at first to run simultaneously several particles starting from the level B_i ; after a while, some of them have evolved badly , the other have well evolved i.e. have succeeded in reaching the threshold B_{i+1}. Bad particles are then moved to the position of the good ones and so on until A 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. Examples of this class of algorithms can be found in [2] with the go with the winners scheme, in [13, 21] in the context of the approximate counting and in [10, 11, 14] in a more general setting.

Nevertheless, in practice the trajectory splitting method may be difficult to apply. For example, the case of the estimation of the probability of a rare event in random dynamical systems is more complex, since the difficulty to find theoretically the optimal B_i. Furthermore, the probability to reach B_i varies generally with the state of entrance in level B_{i-1}. But it is not always the case e.g. for Markovian models (like diffusions).

A mathematical tool well adapted to study this type of algorithms is the Feynman-Kac approach developped in [11]. Asymptotic results are derived, such as LLN, CLT, and Large Deviations principles ; in particular asymptotic variance of the estimator of the rare event probability is given. Non asymptotic results such as uniform Lp mean error bounds and exponential concentration inequalities with respect to the time horizon can be also found in this relevant book. Getting precise confidence intervals is more challenging. Nevertheless, all these algorithms lie on a common base, simpler to analyse and called branching splitting model. In this technique, interactions between particles are avoided and its relative simplicity allow us to derive precise results (Chernoff type bound of the estimator) and to have better knowledge and understanding on splitting models in general. We must precise here that we consider only one dimensional models as introduced in Garvels [15] or in a more refined version : the RESTART method [30, 31].

In the branching splitting technique, make a $\{0, 1\}$ Bernoulli trial to check whether or not the set event B_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 B_2 has occured. This procedure is repeated at each level, until A is reached. If an event level is not reached, neither is A , then we stop the current retrial. Using N independent replications of this procedure, we have then considered $NR_1 \dots R_M$ trials, taking into account for example, that if we have failed to reach a level B_i at the i-th step, the $R_i \dots R_M$ possible retrials have failed. Clearly the particles reproduce and evolve independently.

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

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

where N_A is the total number of trajectories having reached the set A. Considering that this algorithm is represented by N independant Galton-Watson branching processes $(Z_n)_n$, as done in [23], 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 [30, 31], 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 (Lagnoux [23]) suggests to take all the transition probability equal to a constant P_0 and the number 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.

Now, we are interested in the study of the precision of this algorithm by deriving an upper bound to the quantity

$$\mathbb{P}(|\widehat{P} - \mathbb{P}(A)|/\mathbb{P}(A) \geq \alpha) \tag{3}$$

The Chebycheff's bound being too crude, we will use the Chernoff type bound based on the Laplace transfom of the normalized Galton-Watson branching process

$$W_{M+1} := Z_{M+1}/\mathbb{E}(Z_{M+1}).$$

We therefore need to get estimates on the Laplace transform of W_{M+1} which depends on the n-th iterate of a function ψ. In practice, the optimal number of thresholds M is not very large and therefore asymptotic estimates may not be accurate. That optimal number being derived from these bounds, a numerical approach would be unworth too. Hence we want to derive explicit upper and lower bounds for a given number of thresholds. For an asymptotic analysis of the branching splitting model, the reader is referred to Glasserman et al. [16, 17].

In this paper, precise estimates are derived using the following technique : instead of a single bounding function for ψ, several ones are used to obtain sharper upper and lower estimates. These bounding functions are chosen in the low dimensional Lie groups of the homographic and affine functions. The higher the dimension of the Lie group, the more precise is the approximation since the dimension describes the number of parameters to adjust the function with. Unfortunately, we did not find higher dimensional Lie algebras of monotone functions (necessary property to iterate merely the inequalities obtained). The interest of using such functions lies also in the fact that their iterates can be explicitly computed. This technique leads to accurate bounds on the probability (3) (see Proposition 1).

In practice, the adjustment of P_i close to the optimal value may be done during a first phase. The proportion of the cost devoted to this learning part will be the topic of a forthcoming paper (see Section 7). But it soon appears that, even in the case of P_i's close to optimals, the fact that the number of replicas is not an integer destroys rapidly the accuracy of the algorithm : in such a case, one can take R_i equal to the closest integer (k or $k+1$) of the

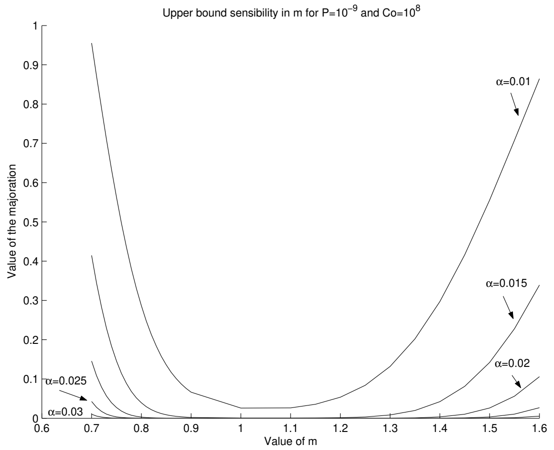


FIG. 1. Sensitivity of the bounds in m for different values of α

optimal value R but whatever the choice we have made, the criticality of the Galton-Watson process will be lost and the loss of precision is significant, see Figure 1 and Tabular 1.

	$\mathbb{P}(A) = 3.5\ 10^{-11}$ and $C = 10^8$		$\mathbb{P}(A) = 5.10^{-11}$ and $C = 10^8$	
	Variance	Laplace	Variance	Laplace
Optimal $m = 1$	0.05080	0.03948	0.04938	0.03835
	Supercritical case : $m = 1.111$		Subcritical case : $m = 0.909$	
Deterministic $R_i = [R] + \mathbf{1}_{R-[R]\geq 0.5}$	0.05702	0.04428	0.05428	0.04215

TAB1. *Length of the* 95% *confidence interval given by the variance and the Laplace transform*

This paper is concerned with the study of different ways to overcome this problem. Lead by [2], we choose at random the sampling number with the hope of improving the simulation. In a first model (Random1), we sample a Bernoulli random variable R_i on $\{k, k+1\}$ for each particle having reached level i started from level i-1 and we decide to adjust $p := \mathbb{P}(R_1 = k)$ such that $m = 1$. A second model (Random2) consists in sampling a random environmental sequence $(R_1, R_2, \ldots, R_M)$ of M i.i.d. Bernoulli random variables R_i on $\{k, k+1\}$ with common parameter p, derived by the same previous optimization approach with an additional constraint (the link between the expectations of R and its inverse). However, this problem is more complex and needs an approximate solution. Results are presented in Proposition 3 and in Tabular 2.

	$\mathbb{P}(A) = 3.5\ 10^{-11}$ and $C = 10^8$		$\mathbb{P}(A) = 5.10^{-11}$ and $C = 10^8$	
	Variance	Laplace	Variance	Laplace
Optimal m=1	0.05080	0.03948	0.04938	0.03835
	Supercritical case : $m = 1.111$		Subcritical case : $m = 0.909$	
Deterministic $R_i = [R] + \mathbf{1}_{R-[R] \geq 0.5}$	0.05702	0.04428	0.05428	0.04215
Random 1	0.05134	0.03990	0.04990	0.03878
Random 2	0.05388	0.04026	0.05235	0.03910

TAB2. *Length of the* 95% *confidence interval given by the variance and the Laplace transform*

Remark that Random 1 provides the closest results from the optimals.

1.2 The results

Using the technique based on low dimensional Lie groups, described in section 1.1, we shall prove the following precise estimates :

Proposition 1 *For a given cost C, there exists a generic constant α_1 such that for $\alpha \leq \alpha_1$,*

$$\mathbb{P}\left(\frac{|\hat{P} - \mathbb{P}(A)|}{\mathbb{P}(A)} \geq \alpha\right) \leq \begin{cases} 2\exp\left\{-\frac{C\alpha^2 P_0}{8(1-P_0)h(P_0)} m^M \left(\frac{m-1}{m^{M+1}-1}\right)^2\right\} & \text{for } m \neq 1 \\ 2\exp\left\{-\frac{C\alpha^2 P_0}{8(1-P_0)h(P_0)} \frac{1}{(M+1)^2}\right\} & \text{for } m = 1 \end{cases}$$

The value of α_1 will be given in Section 3.

Remark 2 *For large value of M, the bound of (3) then behaves like*

$$2\exp\left\{-\frac{C\alpha^2 P_0}{8(1-P_0)h(P_0)}\frac{1}{m^{M+2}}\right\} \quad for \quad m>1$$
$$2\exp\left\{-\frac{C\alpha^2 P_0}{8(1-P_0)h(P_0)}\frac{1}{(M+1)^2}\right\} \quad for \quad m=1$$
$$2\exp\left\{-\frac{C\alpha^2 P_0}{8(1-P_0)h(P_0)}m^M\right\} \quad for \quad m<1$$

With (13) and (10) in Section 2, we obtain the following gaussian bound

$$2\exp\left(-\frac{\alpha^2}{8}\frac{N}{\operatorname{var}(W_{M+1})}\right) \tag{4}$$

Considering R as a random variable in the two ways described in Section 1.1 and using the same type of techniques we get

Proposition 3 *(1)* ***Sampling at each success***
There exists generic constants c_2 and α_2 such that for $\alpha \le \alpha_2$

$$\mathbb{P}\left(\frac{|\hat{P}-\mathbb{P}(A)|}{\mathbb{P}(A)} \ge \alpha\right) \le 2\exp\left\{-c_2\frac{C\ \alpha^2}{h(P_0)}\frac{1}{(M+1)^2}\right\}$$

where c_2 is a constant depending on p and k.

(2) ***Sampling a random environment***
There exists a generic constant α_3 such that for $\alpha \le \alpha_3$

$$\mathbb{P}\left(\frac{|\hat{P}-\mathbb{P}(A)|}{\mathbb{P}(A)} \ge \alpha\right) \le 2\mathbb{E}\left(\exp\left\{-\frac{\alpha^2}{8(1-P_0)}\frac{1}{\sum_{i=0}^{M}\xi_0^{-1}\dots\xi_i^{-1}}\right\}\right)^{\widetilde{C}}$$

where $\xi_0 = P_0$, $\xi_i = R_i P_0$ for $i = 1\dots M$, $\widetilde{C} = \frac{C}{h(P_0)}\frac{\mathbb{E}(\xi)-1}{\mathbb{E}(\xi)^{M+1}-1}$.

The paper is organized as follows. Section 2 recall quickly the general settings of the branching splitting model. We obtain, in Section 3, estimates on (3) and their sensitivity in m. In Sections 4 and 5, the two random models are studied and we derive precise estimates on (3). Section 6 presents a numerical illustration. Finally in Section 7, we conclude and discuss the merits of this approach and potential directions for further researches.

2 Settings of the branching splitting model

As said in Introduction and following [23], we consider N independent Galton-Watson branching processes $(Z_n^{(i)})_{n\geq 0}$, $i = 1,\dots,N$ 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 B_n. Then, letting R_i the sampling number at level i,

$$\widehat{P} := \frac{1}{N}\sum_{i=1}^{N}\widetilde{P}_i, \quad \text{where} \quad \widetilde{P}_i = \frac{Z_{M+1}^{(i)}}{R_1\dots R_M}. \tag{5}$$

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 recurrence relation

$$Z_{n+1} = \sum_{j=1}^{Z_n} X_n^{(j)} \tag{6}$$

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 $n = 0$. Let introduce the following quantities

$$m_0 = P_1, \qquad m_n = R_n P_{n+1}, \quad n = 1,\dots,M+1,$$

which are the mean number of particles success at each level. Then

$$\widetilde{P} = \frac{Z_{M+1}}{R_1\dots R_M} = \mathbb{P}(A)\frac{Z_{M+1}}{m_0\dots m_M} \tag{7}$$

According to (6), we obtain that $\mathbb{E}(Z_{n+1}|Z_n) = m_n Z_n$, so using this relation repeatedly, we find that $\mathbb{E}(Z_{n+1}) = \prod_{i=0}^{n} m_i$. Introducing the random variable $W_n = Z_n/\mathbb{E}(Z_n)$ gives the new expression

$$\widetilde{P} = \mathbb{P}(A)W_{M+1} \tag{8}$$

from which, we easily deduce that the estimate $\widehat{P}$ is unbiased. Integrating N, the variance of $\widehat{P}$ is given by

$$\operatorname{var}(\widehat{P}) = \frac{\mathbb{P}(A)^2}{N}\sum_{i=0}^{M}\frac{1-P_{i+1}}{m_0\dots m_i}. \tag{9}$$

The minimization of the variance of $\mathbb{P}(A)$ for a given budget C defined by

$$C = N\left[h(P_1) + \sum_{i=1}^{M} h(P_{i+1})m_0\dots m_{i-1}R_i\right] \tag{10}$$

leads to the optimal parameters of the algorithm given by [23]

$$\begin{cases} P_i &= P_0, \quad i = 1, \dots, M+1, \\ m_i &= 1, \quad i = 1, \dots, M, \\ N &= C/[(M+1)h(P_0)], \\ M &= \lfloor \log \mathbb{P}(A)/y_0 \rfloor, \end{cases}$$

where y_0 is the solution of some equation which depends on the unit cost function h. It means that the branching processes are critical Galton Watson processes. The reader is referred to Harris [19], Lyons [24] and Athreya and Ney [5] for more details on Galton-Watson and branching processes.

Remind the goal of this paper is to study the precision of the algorithm by deriving an upper bound to (3) that decomposes in two probabilities itselves bounded by

$$\mathbb{P}\left(\widehat{P} \geq (1+\alpha)\mathbb{P}(A)\right) \leq \exp\left\{N \inf_{u>0} \left[\log \mathbb{E}\left(e^{uW_{M+1}}\right) - u(1+\alpha)\right]\right\}, \qquad (11)$$

$$\mathbb{P}\left(\widehat{P} \leq (1-\alpha)\mathbb{P}(A)\right) \leq \exp\left\{N \inf_{u<0} \left[\log \mathbb{E}\left(e^{uW_{M+1}}\right) - u(1-\alpha)\right]\right\}. \qquad (12)$$

Thus precise estimates of the Laplace transform of W_{M+1} need to be derived. In the following, we consider our model with all the P_i's equal to a constant P_0 and all the R_i's equal to R ; but $m = RP_0$ could be different from the optimal value 1. From (9) and (10), remark

$$\operatorname{var}(W_{M+1}) := \sigma^2_{M+1} = \begin{cases} \left(\frac{1}{P_0} - 1\right) \frac{m^{M+1}-1}{(m-1)m^M} & \text{for } \ m \neq 1, \\ \left(\frac{1}{P_0} - 1\right)(M+1) & \text{for } \ m = 1, \end{cases} \qquad (13)$$

and

$$C = \begin{cases} Nh(P_0)\frac{m^{M+1}-1}{m-1} & \text{for } \ m \neq 1, \\ Nh(P_0)(M+1) & \text{for } \ m = 1. \end{cases} \qquad (14)$$

3 First model : taking R deterministic with $m \neq 1$

3.1 The model

Throughout this section, we consider the branching process associated to the splitting algorithm in the case R deterministic with $m \neq 1$ and we aim at deriving relevant confidence intervals.

Note the generating function of Z_{M+1} is $P_0 f^{oM}(s) + 1 - P_0$ where $f(s) = (P_0 s + 1 - P_0)^R$. To lighten notation, we denote the n-th functional iterate of a function f by f_n.

Letting $\psi(u) := \log(f(e^u)) = R\log(P_0 e^u + 1 - P_0)$ and $h(u) = \exp(u)$, we observe that $\psi = h^{-1} o f o h$; thus $\psi_{M+1} = h^{-1} o f_{M+1} o h$ and the Laplace transform of W_{M+1} becomes

$$\mathbb{E}\left(e^{uW_{M+1}}\right) = P_0 f_M\left(e^{\frac{u}{P_0 m^M}}\right) + 1 - P_0 = \exp\left\{\frac{1}{R}\psi_{M+1}\left(\frac{u}{P_0 m^M}\right)\right\} \tag{15}$$

As a consequence, (11) and (12) yield to

$$\mathbb{P}\left(\frac{|\hat{P} - \mathbb{P}(A)|}{\mathbb{P}(A)} \geq \alpha\right) \leq \exp\left\{N \inf_{u>0} F^+_{M+1}(u)\right\} + \exp\left\{N \inf_{u<0} F^-_{M+1}(u)\right\}$$

where

$$\begin{cases} F^+_{M+1}(u) := \frac{1}{R}\psi_{M+1}\left(\frac{u}{P_0 m^M}\right) - (1+\alpha)\,u \\ F^-_{M+1}(u) := \frac{1}{R}\psi_{M+1}\left(\frac{u}{P_0 m^M}\right) - (1-\alpha)\,u \end{cases}$$

Remark 4 *Note that for $u \leq 0$, $\mathbb{E}\left(e^{uW_n}\right) \leq 1$ and so $\mathbb{E}\left(e^{uW}\right) \leq 1$. What happens for $u \geq 0$? We have*

$$\psi_{n+1}\left(\frac{u}{P_0 m^n}\right) = R\log\mathbb{E}\left(e^{uW_{n+1}}\right)$$

As a consequence, information on $\psi_{n+1}\left(\frac{u}{P_0 m^n}\right)$ provides us results on the exponential integrability of W_{n+1} and W.

3.2 The Laplace transform of W_{n+1}

3.2.1 Introduction

Wanting to obtain precise bounds on the Laplace transform of W_{n+1} given by (15) (M being replaced by n), we are interested in the behavior of ψ_{n+1}. But since ψ_{n+1} has no explicit expression, we are brought to estimate ψ in a first step and then to iterate these estimates.

Heuristic for estimating ψ : instead of using a single function to bound ψ, we use several bounding functions to obtain sharper upper and lower estimates. Near the origin, we will use homographic functions and elsewhere affine functions.

For example, for $u \geq 0$ and $m > 1$, near the origin a good homographic function (same value at 0, same first and second derivatives at 0) gives a sharp upper bound. For large u, we simply use the asymptotic direction. We get

$$\psi(u) \leq \begin{cases} mu/\left(1 - \frac{1-P_0}{2}u\right) & \text{for } 0 \leq u \leq u_0 \\ Ru \text{ for } u_0 \leq u \end{cases}$$

for u_0 such that $mu_0/\left(1 - \frac{1-P_0}{2}u_0\right) = Ru_0$.

Let us introduce $h(u;x,y) = \frac{xu}{1-yu}$ and $g(u;x,y) = xu + y$. More generally, we obtain upper and lower estimates of the following form

$$\begin{cases} h(u;x,y) & \text{for } u \text{ close to } 0 \\ g(u;x,y) & \text{otherwise} \end{cases}$$

We choose these bounding functions in the low dimensional Lie groups of the homographic and affine functions. The higher the dimension of the Lie group, the more precise is the approximation since the dimension describes the number of parameters to adjust the function with (for example, the k first derivatives at a given point). Unfortunately, we did not found higher dimensional Lie algebras of functions having the monotonicity assumption required to iterate merely the inequalities obtained for ψ. The interest of using such functions lies also in the fact that their iterates can be explicitly computed. Obviously

$$h_n(u;x,y) = \frac{x^n u}{1 - yu\frac{x^n-1}{x-1}} \quad \text{and} \quad g_n(u;x,y) = x^n u + y\frac{x^n - 1}{x - 1}$$

The problem is that the set formed by the homographic functions and the affine functions is not a group; so a crucial point will be to determine the number of iterations corresponding to a change of group.

Heuristic for estimating ψ_n : Since the estimates of ψ have two expressions, we will obtain estimates of ψ_n with three functions. Indeed, starting from u close to 0, we iterate h, starting from u far from 0, we iterate the affine function g. Nevertheless when u starts from a midle value, we shall encounter a change of regime as described below.

More precisely, in the case when $m > 1$ and $u > 0$ e.g., 0 is a repulsive fix point (see next section). So starting from $u \geq 0$, $\psi_n(u)$ diverges while n goes to ∞ and there are different regimes for the upper estimates :
- if $u \geq u_0$, we simply iterate $u \mapsto Ru$ and the n-fold convolution of ψ is lower than the n-fold convolution of Ru,

– if $u \leq u_0$, we start by iterating $h\left(.; m, \frac{1-P_0}{2}\right)$ and if fortunately after n iterations $h_n\left(u; m, \frac{1-P_0}{2}\right)$ is still lower than u_0 then the n-fold convolution of ψ is lower than the n-fold convolution of h. In the other case, after k iterations, $h_k\left(u; m, \frac{1-P_0}{2}\right)$ is greater than u_0 then we continue iterating $u \mapsto Ru$. The value k_0 appearing in the following lemma corresponds to that change of regime.

Finally, the estimates of ψ_n reflects that change of regime and have the following form

$$\begin{cases} h_n(u; x, y) \text{ for } u \text{ close to } 0 \\ g_n(u; x, y) \text{ for } u \text{ far from } 0 \\ au^\alpha \text{ otherwise} \end{cases}$$

3.2.2 Behavior and estimates of ψ for $m > 1$

We plot in Figure 2 ψ in the case $m > 1$. Note that ψ has two fix points : 0 which is repulsive and some $\tilde{q} < 0$ which is attractive (s.t. $\psi(\tilde{q}) = \tilde{q}$).

Proceeding as explained in the previous section, straightforward studies of

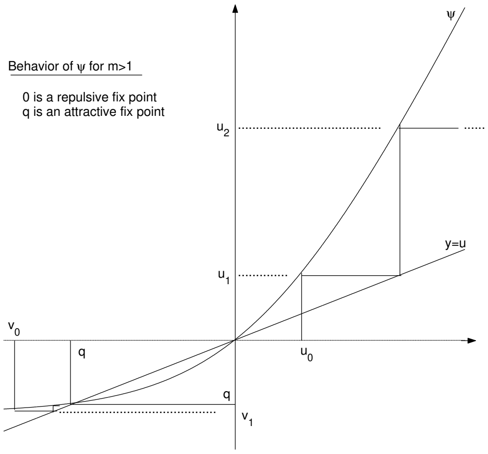


FIG. 2. Behavior of ψ for $m > 1$

functions leads to the following lemmas.

Lemma 3.1 *Estimates of ψ for $u \geq 0$ and $m > 1$*

$$\psi(u) \leq \begin{cases} h\left(u; m, \frac{1-P_0}{2}\right) & \text{for } 0 \leq u \leq u_0 \\ Ru & \text{for } u \geq u_0 \end{cases}$$

$$\psi(u) \geq \begin{cases} mu \quad \textit{for} \quad u \leq v_0 \\ g\left(u; R, R\log\left(P_0\right)\right) \quad \textit{for} \quad u \geq v_0 \end{cases}$$

where $u_0 := 2$ *and* $v_0 := -\log\left(P_0\right)/\left(1-P_0\right) \geq 0$.

Lemma 3.2 *Estimates of* ψ *for* $u \leq 0$ *and* $m > 1$. *Let*

- b *such that* $h\left(b; m, 1-P_0\right) < b$,
- $d := [\tilde{q} - h\left(b; m, 1-P_0\right)]/(\tilde{q}-b)$,
- $m_{\tilde{q}} := \psi'\left(\tilde{q}\right) = me^{\tilde{q}\left(1-\frac{1}{R}\right)}$,
- $u_{-\infty} := R\log\left(1-P_0\right)$,
- $v_1 := arg_u\{h\left(u; m, 1-P_0\right) = g\left(u; m_{\tilde{q}}, \tilde{q}(1-m_{\tilde{q}})\right)\}$,
- $v_2 := arg_u\{g\left(u; m_{\tilde{q}}, \tilde{q}(1-m_{\tilde{q}})\right) = u_{-\infty}\} = \tilde{q} + (u_{-\infty} - \tilde{q})/m_{\tilde{q}}$.

$$\psi(u) \leq \begin{cases} h\left(b; m, 1-P_0\right) \quad \textit{for} \quad b \leq u \leq 0 \\ g\left(u; d, \tilde{q}(1-d)\right) \quad \textit{for} \quad \tilde{q} \leq u \leq b \\ \tilde{q} + h\left(u-\tilde{q}; m_{\tilde{q}}, (1-P_0)e^{-\frac{\tilde{q}}{R}}\right) \quad \textit{for} \quad u_2 \leq u \leq \tilde{q} \quad \textit{for some} \quad u_2 \leq 0 \end{cases}$$

$$\psi(u) \geq \begin{cases} h(u; m, \frac{1-P_0}{2}) \quad \textit{for} \quad v_1 \leq u \leq 0 \\ g\left(u; m_{\tilde{q}}, \tilde{q}(1-m_{\tilde{q}})\right) \quad \textit{for} \quad v_2 \leq u \leq v_1 \\ u_{-\infty} \quad \textit{for} \quad u \leq v_2 \end{cases}$$

3.2.3 Estimates of ψ_n for $m > 1$

Theorem 3.1 *Estimates of* ψ_n *for* $u \geq 0$ *and* $m > 1$ *Let* $u_{\star,n} := h_n^{-1}\left(u_0; m, \frac{1-P_0}{2}\right) = 2/[m^n(1+\sigma_n^2/R)]$ *and* $\gamma := \log R/\log m > 1$.
Then there exists δ_i *for* $i = 0, 1$ *such that*

$$\psi_n(u) \leq \begin{cases} h_n(u; m, \frac{1-P_0}{2}) \quad \textit{for} \quad 0 \leq u \leq u_{\star,n} \\ u_0 R^{n-\delta_0}\left(\frac{u}{1+\frac{1-P_0}{2(m-1)}u}\right)^{\gamma} \quad \textit{for} \quad u_{\star,n} \leq u \leq u_0 \\ R^n u \quad \textit{for} \quad u_0 \leq u \end{cases}$$

$$\psi_n(u) \geq \begin{cases} m^n u \quad \textit{for} \quad 0 \leq u \leq \frac{v_0}{m^n} \\ R^{n-\delta_1}u^{\gamma}\left(v_0 - q_g\right) \quad \textit{for} \quad \frac{v_0}{m^n} \leq u \leq v_0 \\ g_n\left(u; R, R\log P_0\right) \quad \textit{for} \quad v_0 \leq u \end{cases}$$

Proof Upper bound
During the proof we write h instead of $h(.;m,\frac{1-P_0}{2})$.
If $\mathbf{u} \leq \mathbf{u}_{\star,\mathbf{n}}$,
-since $m>1$, $u \leq h(u)$ and since h is increasing, $u \leq h(u) \leq \ldots \leq h_n(u)$.
-since $u \leq u_{\star,n}$, $h_n(u) \leq u_0$ then

$$\psi_n(u) = \psi_{n-1}(\psi(u)) \leq \psi_{n-1}(h(u)) \leq \ldots \leq h_n(u) = \frac{m^n u}{1 - \frac{1-P_0}{2} u \frac{m^n-1}{m-1}}$$

If $\mathbf{u} \geq \mathbf{u_0}$, since $R \geq 1$, $u_0 \leq u \leq Ru \leq \ldots \leq R^n u$ and so $\psi_n(u) \leq R^n u$

If $\mathbf{u} \in [\mathbf{u}_{\star,\mathbf{n}}, \mathbf{u_0}]$, since $u \leq u_0$, we start by iterating h and since $u \leq h(u)$ and $u \geq h_n^{-1}(u_0)$, the iterates $h_k(u)$ finally become greater than u_0 and then we iterate $u \mapsto Ru$. Let k_0 such that $h_{k_0}(u) \leq u_0$ and $h_{k_0+1}(u) > u_0$. k_0 corresponds to the change of regime and

$$k_0 := \log\left(\frac{u_0}{1+\frac{1-P_0}{2(m-1)}u_0}\right)/\log m - \log\left(\frac{u}{1+\frac{1-P_0}{2(m-1)}u}\right)/\log m$$

Letting $\delta_0 = \log\left(\frac{u_0}{1+\frac{1-P_0}{2(m-1)}u_0}\right)/\log m$, we get

$$\begin{aligned}\psi_n(u) &= \psi_{n-k_0}(\psi_{k_0}(u)) \leq R^{n-k_0} h_{k_0}(u) \leq R^{n-k_0} u_0 \\ &\leq \ldots \leq u_0 R^{n-\delta_0}\left(\frac{u}{1+\frac{1-P_0}{2(m-1)}u}\right)^{\gamma}\end{aligned}$$

□

A similar behavior is observed in the other cases thus we will omit their proofs.

Theorem 3.1 allows us to state about the exponential integrabibilty of W, the limit of W_n as $n \to +\infty$:

Corollary 5 *The random variable $uW^{\gamma'}$ is exponentially integrable, independently of the value of u, where $\gamma' = \frac{\gamma}{\gamma-1} \geq 1$ i.e.*

$$\mathbb{E}\left(e^{uW^{\gamma'}}\right) < +\infty \quad \textit{for all} \quad u \in \mathbb{R}$$

Theorem 3.2 *Estimates of ψ_n for $u \leq 0$ and $m>1$. Let*
– $u_{\star\star,n} := h_n^{-1}(b;m,1-P_0) = b/[m^n(1+\sigma_n^2 b/R)]$,
– $v_{\star,n} := h_n^{-1}\left(v_1;m,\frac{1-P_0}{2}\right) = v_1/[m^n(1+\sigma_n^2 v_1/(2R))]$,
– $v_{\star\star,n} := g_n^{-1}(v_2;m_{\tilde{q}},\tilde{q}(1-m_{\tilde{q}})) = \tilde{q} - (\tilde{q}-v_2)/m^n$,

– $\eta := \log d / \log m$ *and* $\mu := \log m / \log m_{\tilde{q}}$.

Then there exists δ_i *for* $i = 2, 3$ *such that*

$$\psi_n(u) \leq \begin{cases} h_n(u; m, 1-P_0) \quad \textit{for} \quad u_{\star\star,n} \leq u \leq 0 \\ \tilde{q} + d^{n-\delta_2}\left(\frac{u}{1+\frac{1-P_0}{m-1}}\right)^{\eta}(b-\tilde{q}) \quad \textit{for} \quad b \leq u \leq u_{\star\star,n} \\ g_n(u; d, \tilde{q}(1-d)) \quad \textit{for} \quad \tilde{q} \leq u \leq b \\ \tilde{q} + h_n\left(u - \tilde{q}; m_{\tilde{q}}, (1-P_0)e^{-\frac{\tilde{q}}{R}}\right) \quad \textit{for} \quad u_2 \leq u \leq \tilde{q} \end{cases}$$

$$\psi_n(u) \geq \begin{cases} h_n(u; m, \frac{1-P_0}{2}) \quad \textit{for} \quad v_{\star,n} \leq u \leq 0 \\ \tilde{q} + m_{\tilde{q}}^{n-\delta_3}\left(\frac{u}{1+\frac{1-P_0}{2(m-1)}u}\right)^{-\mu}(v_1-\tilde{q}) \quad \textit{for} \quad v_1 \leq v_{\star,n} \\ g_n(u; m_{\tilde{q}}, \tilde{q}(1-m_{\tilde{q}})) \quad \textit{for} \quad v_{\star\star,n} \leq u \leq v_1 \\ u_{-\infty} \quad \textit{for} \quad u \leq v_{\star\star,n} \end{cases}$$

3.2.4 Estimates of ψ_n for $m < 1$

Proceeding in the same way, we obtain the same kind of inequalities.

3.3 Confidence intervals

Theorem 3.3 *Bounds of* $\mathbb{P}\left(\frac{|\hat{P}-\mathbb{P}(A)|}{\mathbb{P}(A)} \geq \alpha\right)$

In both cases, for α *small enough and* $m \neq 1$,

$$\mathbb{P}\left(\frac{|\hat{P}-\mathbb{P}(A)|}{\mathbb{P}(A)} \geq \alpha\right) \leq h_+ + h_-$$

where

$$\begin{cases} h_+ := \exp\left\{-\frac{2C}{h(P_0)(1-P_0)R}m^{M+1}\left(\frac{1-m}{1-m^{M+1}}\right)^2\left(\sqrt{1+\alpha}-1\right)^2\right\} \\ h_- := \exp\left\{-\frac{C}{h(P_0)(1-P_0)R}m^{M+1}\left(\frac{1-m}{1-m^{M+1}}\right)^2\left(1-\sqrt{1-\alpha}\right)^2\right\} \end{cases}$$

Proof All the cases behave similarly : so we just treat the case $m > 1$ and $u \geq 0$ for example. We have $F^+_{M+1}(u) := \frac{1}{R}\psi_{M+1}\left(\frac{u}{P_0 m^M}\right) - (1+\alpha)u$. Then

$$\inf_{u>0} F^+_{M+1}(u) = \inf\left\{\inf_{u\in[0,u_{\star,M+1}]} F^+_{M+1}(u), \inf_{u\in[u_{\star,M+1},u_0]} F^+_{M+1}(u), \inf_{u>u_0} F^+_{M+1}(u)\right\}$$

For $\mathbf{u} \in [\mathbf{0}, \mathbf{u}_{\star,\mathbf{M+1}}]$,

$$F^{+}_{M+1}(u) \leq \frac{u}{1-\sigma^2_{M+1}u/2} - u(1+\alpha) := \phi_{M+1}(u)$$

The cancellation of the first derivative of ϕ_{M+1} gives us the minimum of ϕ_{M+1} :

$$u_s = \frac{2}{\sigma^2_{M+1}}\left(1 - \frac{1}{\sqrt{1+\alpha}}\right)$$

which is lower than $u_{\star,M+1}$ for α small enough i.e.

$$\alpha \leq \left(1 - u_{\star,M+1}\sigma^2_{M+1}/2\right)^{-2} - 1$$

For $\mathbf{u} \in [\mathbf{u}_{\star,\mathbf{M+1}}, \mathbf{u_0}]$, $F^{+}_{M+1}(u) \leq R^{M-\delta_0}u_0\left(\frac{u}{1+\frac{1-P_0}{2(m-1)}\frac{u_{\star,M+1}}{P_0 m^M}}\right)^{\gamma} - (1+\alpha)u := \phi_{M+1}(u)$. The cancellation of the first derivative of ϕ_{M+1} gives us the minimum of ϕ_{M+1} :

$$\tilde{u}_s = \left\{\frac{(1+\alpha)}{\gamma u_0 R^{M-\delta_0}}\left(P_0 m^M + \frac{1-P_0}{2(m-1)}u_{\star,M+1}\right)^{\gamma}\right\}^{\frac{1}{\gamma-1}}$$

which is lower than $u_{\star,M+1}$ for

$$\alpha \leq \frac{\gamma u_0 R^{M-\delta_0}}{u_{\star,M+1}}\left(\frac{u_{\star,M+1}}{P_0 m^M + \frac{1-P_0}{2(m-1)}u_{\star,M+1}}\right)^{\gamma} - 1$$

And so the solution leaves the interval.

For $\mathbf{u} \geq \mathbf{u_0}$,

$$F^{+}_{M+1}(u) \leq R^M u/\left(P_0 m^M\right) - u(1+\alpha) = u\left[1/P_0^{M+1} - (1+\alpha)\right],$$

and $1/P_0^{M+1} - (1+\alpha)$ is positive for M great enough; so the minimum is atteined in u_0 and the solution leaves the interval.
Finally, $\mathbb{P}\left(\hat{P} \geq P(1+\alpha)\right)$ is bounded by

$$\exp\left\{\frac{N}{R}F^{+}_{M+1}(u_s)\right\} = \exp\left\{-\frac{N}{R}\frac{1}{1-P_0}m^{M+1}\frac{1-m}{1-m^{M+1}}\left(\sqrt{1+\alpha}-1\right)^2\right\}$$

and we get the result from (14). □

Theorem 3.4 *For $m = 1$ and α small enough, we have*

$$\mathbb{P}\left(\frac{|\hat{P} - \mathbb{P}(A)|}{\mathbb{P}(A)} \geq \alpha\right) \leq g_+ + g_-$$

where

$$\begin{cases} g_+ := \exp\left\{-\frac{2C}{h(P_0)(1-P_0)R}\frac{1}{(M+1)^2}\left(\sqrt{1+\alpha}-1\right)^2\right\} \\ g_- := \exp\left\{-\frac{C}{h(P_0)(1-P_0)R}\frac{1}{(M+1)^2}\left(1-\sqrt{1-\alpha}\right)^2\right\} \end{cases}$$

We check that this bound is better than the one for $m \neq 1$.

The next corollary follows from Theorem 3.3 and provides us an exact upper bound of $\mathbb{P}\left(|\hat{P} - \mathbb{P}(A)|/\mathbb{P}(A) \geq \alpha\right)$ involving the variance of W_{M+1} that we might compare to (4), the approximate one obtained in the Introduction.

Corollary 6 *For α such as in Theorem 3.3,*

$$\mathbb{P}\left(\frac{|\hat{P} - \mathbb{P}(A)|}{\mathbb{P}(A)} \geq \alpha\right) \leq 2\exp\left\{-\frac{\alpha^2}{4}\frac{N}{var(W_{M+1})}\left(1-\frac{\alpha}{2}\right)\right\}$$

Proof By Theorem 3.3, the definition of N and (13), we deduce that $\mathbb{P}\left(|\hat{P} - \mathbb{P}(A)|/\mathbb{P}(A) \geq \alpha\right)$ is bounded by

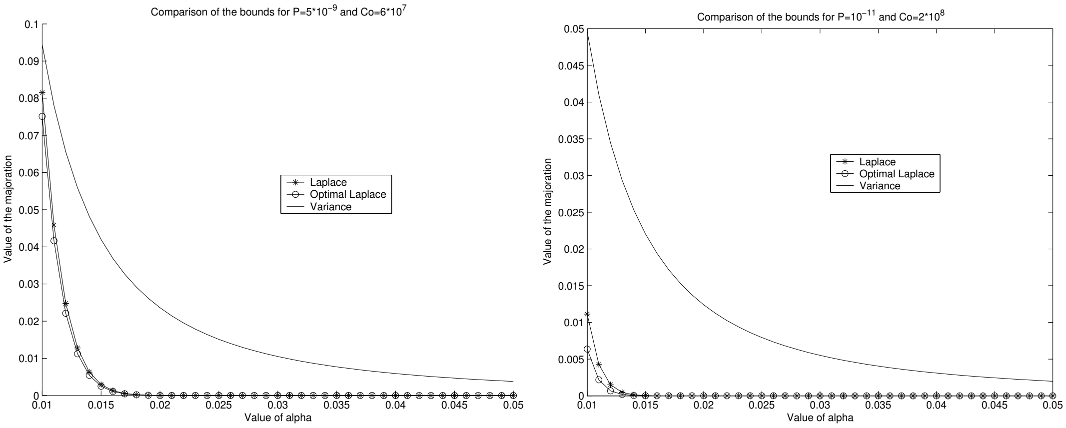
$$\exp\left\{-\frac{2N}{\mathrm{var}(W_{M+1})}\left(\sqrt{1+\alpha}-1\right)^2\right\} + \exp\left\{-\frac{N}{\mathrm{var}(W_{M+1})}\left(1-\sqrt{1-\alpha}\right)^2\right\}$$

But for $\alpha \leq 2$ and since $\sqrt{1+\alpha} \leq 1 + \frac{\alpha}{2}$ and $\sqrt{1-\alpha} \leq 1 - \frac{\alpha}{2}$,

$$\begin{aligned} \left(\sqrt{1+\alpha}-1\right)^2 &= \frac{\alpha^2}{2+\alpha+2\sqrt{1+\alpha}} \geq \frac{\alpha^2}{4}\frac{1}{1+\frac{\alpha}{2}} \geq \frac{\alpha^2}{4}\left(1-\frac{\alpha}{2}\right) \\ \left(1-\sqrt{1-\alpha}\right)^2 &= \frac{\alpha^2}{2-\alpha+2\sqrt{1-\alpha}} \geq \frac{\alpha^2}{4}\left(1+\frac{\alpha}{2}\right) \geq \frac{\alpha^2}{4}\left(1-\frac{\alpha}{2}\right) \end{aligned}$$

that leads to the corollary. □

Proposition 1 and Remark 2 now follow from the previous theorems, using the same kind of argument as in Corollary 6.



1. Supercritical case   2. Subcritical case

FIG. 3. Upper estimates for $m \neq 1$

We plot in Figure 3 the different bounds using :
– the variance,
– the Laplace transform in the optimal case : $m = 1$,
– the Laplace transform in the other case : $m \neq 1$,
in two different cases
– for $\mathbb{P}(A) = 5\ 10^{-9}$ and $C = 6\ 10^7$, then m=1.055 (supercritical case),
– for $\mathbb{P}(A) = 10^{-11}$ and $C = 2\ 10^8$, then m=0.924 (subcritical case).

4 Second model : sampling at each success

4.1 The model

Let $k \in \mathbb{N}$ such that $k = \lfloor 1/P_0 \rfloor$. We decide here to allow R to vary. More precisely, each time a particle reach a higher level, we generate a realization of a Bernoulli random variable R whose parameter $p = 1 - q$ will be derived by the same optimization scheme as used previously. Let $\xi = RP_0$.

Here the estimator chosen is given by :

$$\hat{P} = \frac{\mathbb{P}(A)}{N} \sum_{i=1}^{N} W_{M+1}^i = \frac{1}{N} \sum_{i=1}^{N} \frac{Z_{M+1}^i}{\mathbb{E}(Z_{M+1})} = \frac{\mathbb{P}(A)}{N} \sum_{i=1}^{N} \frac{Z_{M+1}^i}{P_0 \mathbb{E}(\xi)^M}$$

4.2 The variance and its optimization

4.2.1 Study of the variance

Proposition 7 *$\hat{P}$ is trivially an unbiased estimator and its variance is given by*

$$var(\hat{P}) = \frac{\mathbb{P}(A)^2}{N}\left[\left(\frac{1}{P_0}-1\right)\sum_{i=0}^{M}\frac{1}{\mathbb{E}(\xi)^i} + \frac{var(\xi)}{P_0\mathbb{E}(\xi)}\sum_{i=1}^{M}\frac{1}{\mathbb{E}(\xi)^i}\right] \tag{16}$$

Proof First of all, $\mathrm{var}(\hat{P}) = \frac{\mathbb{P}(A)^2}{N}\frac{1}{[P_0\mathbb{E}(\xi)^M]^2}\mathrm{var}(Z_{M+1})$. The calculation of $\mathrm{var}(\hat{P})$ then amounts to the calculation of $\mathrm{var}(Z_{M+1})$.

Note that for all X random variable and $\mathcal{F}$ filtration,

$$\mathrm{var}(X) = \mathrm{var}(\mathbb{E}(X|\mathcal{F})) + \mathbb{E}(\mathrm{var}(X|\mathcal{F})) \tag{17}$$

Applying (17) to $X = Z_{M+1}$ and $\mathcal{F} = \sigma(Z_M)$ and then to $X = X_M$ and $\mathcal{F} = \sigma(R_M)$, and since $X_k^i \sim Bin(R_k, P_0)$, we get

$$\begin{aligned}
\mathrm{var}(Z_{M+1}) &= \mathrm{var}(\sum_{i=1}^{Z_M} X_M^i) = \mathrm{var}(Z_M\mathbb{E}(X_M|Z_M)) + \mathbb{E}(Z_M\,\mathrm{var}(X_M|Z_M))\\
&= \mathbb{E}(X_M)^2\,\mathrm{var}(Z_M) + \mathrm{var}(X_M)\mathbb{E}(Z_M)\\
&= \mathbb{E}(\mathbb{E}(X_M|R_M))^2\,\mathrm{var}(Z_M) + \mathrm{var}(\mathbb{E}(X_M|R_M)\mathbb{E}(Z_M) + \mathbb{E}(\mathrm{var}(X_M|R_M))\mathbb{E}(Z_M)\\
&= \mathbb{E}(\xi)^2\,\mathrm{var}(Z_M) + \mathrm{var}(\xi)\mathbb{E}(Z_M) + (1-P_0)\mathbb{E}(\xi)\mathbb{E}(Z_M)
\end{aligned}$$

By a recurrent descent,

$$\mathrm{var}(Z_{M+1}) = \mathbb{E}(\xi)^{2M}\,\mathrm{var}(Z_1) + \mathrm{var}(\xi)\sum_{k=1}^{M}\mathbb{E}(Z_k)\mathbb{E}(\xi)^{2(M-k)} + (1-P_0)\mathbb{E}(\xi)\sum_{k=1}^{M}\mathbb{E}(Z_k)\mathbb{E}(\xi)^{2(M-k)}$$

But $\mathrm{var}(Z_1) = P_0(1-P_0)$ and $\mathbb{E}(Z_k) = P_0\mathbb{E}(\xi)^{k-1}$, and so we finally get the result.□

4.2.2 Optimization of the parameters

As done in Lagnoux [23], an optimal algorithm is chosen via the minimization of the variance of $\hat{P}$ for a given budget C, keeping the optimal values for M and P_0. The (average) cost is now

$$C = Nh(P_0)\sum_{i=0}^{M}\mathbb{E}(\xi)^i + NP_0h(1)\sum_{i=0}^{M-1}\mathbb{E}(\xi)^i \tag{18}$$

Neglecting the cost introduced by the generation of the random splitting numbers at each success, we assume, in the following,

$$C = Nh(P_0)\sum_{i=0}^{M}\mathbb{E}(\xi)^i, \qquad (h(1) \ll h(P_0)). \tag{19}$$

Suppose $\mathbb{E}(\xi) \neq 1$, then the variance can be rewritten as

$$\begin{aligned}\mathrm{var}(\hat{P}) &= \frac{\mathbb{P}(A)^2}{N}\left[\left(\frac{1}{P_0} - 1 + \frac{\mathrm{var}(\xi)}{P_0\mathbb{E}(\xi)}\right)\sum_{i=0}^{M}\frac{1}{\mathbb{E}(\xi)^i} - \frac{\mathrm{var}(\xi)}{P_0\mathbb{E}(\xi)}\right]\\ &= \frac{\mathbb{P}(A)^2}{N}\left[\left(\frac{1}{P_0} - 1 + \frac{\mathrm{var}(\xi)}{P_0\mathbb{E}(\xi)}\right)\frac{1}{\mathbb{E}(\xi)^M}\frac{\mathbb{E}(\xi)^{M+1}-1}{\mathbb{E}(\xi)-1} - \frac{\mathrm{var}(\xi)}{P_0\mathbb{E}(\xi)}\right]\end{aligned}$$

and the cost has the following form

$$C = Nh(P_0)\frac{\mathbb{E}(\xi)^{M+1}-1}{\mathbb{E}(\xi)-1}$$

The optimal value of N is given by the cost

$$N = \frac{C}{h(P_0)}\frac{\mathbb{E}(\xi)-1}{\mathbb{E}(\xi)^{M+1}-1}$$

and we have to minimize the expression

$$\mathrm{var}(\hat{P}) = \frac{\mathbb{P}(A)^2}{CP_0}h(P_0)\frac{\mathbb{E}(\xi)^{M+1}-1}{\mathbb{E}(\xi)-1}\left[\left(1 - P_0 + \frac{\mathrm{var}(\xi)}{\mathbb{E}(\xi)}\right)\frac{1}{\mathbb{E}(\xi)^M}\frac{\mathbb{E}(\xi)^{M+1}-1}{\mathbb{E}(\xi)-1} - \frac{\mathrm{var}(\xi)}{\mathbb{E}(\xi)}\right]$$

whose principal term is

$$\frac{\mathbb{P}(A)^2}{CP_0}h(P_0)\frac{1}{\mathbb{E}(\xi)^M}\left[\frac{\mathbb{E}(\xi)^{M+1}-1}{\mathbb{E}(\xi)-1}\right]^2$$

which is minimal for $\mathbb{E}(\xi) = 1$.

4.2.3 Confidence interval

From now on, we take $m = f'(1) = \mathbb{E}(\xi) = P_0(k+q) = 1$ which is equivalent to $p = 1 - q = 1 - \delta$. Here the generating function of the reproduction law is given by

$$f(s) = (P_0e^u + 1 - P_0)^k(P_0qe^u + 1 - P_0q) := \exp\{\psi(\log s)\}$$

Remark 8 *That form reflects the fact that R is randomly chosen between k and $k+1$ following the parameter q.*

Since $\log(P_0 s + 1 - P_0) \leq \frac{1}{k+q}\log f(s)$ and $m = 1$,

$$\mathbb{E}\left(e^{uW_{M+1}}\right) \leq \exp\left\{\frac{1}{k+q}\log f_{M+1}(e^{\frac{u}{P_0}})\right\} = \exp\left\{\frac{1}{k+q}\psi_{M+1}\left(\frac{u}{P_0}\right)\right\}$$

Using again the Chernoff's bounding method,

$$\mathbb{P}\left(\frac{|\hat{P} - \mathbb{P}(A)|}{\mathbb{P}(A)} \geq \alpha\right) = \exp\left\{N \inf_{u>0} F^{+}_{M+1}(u)\right\} + \exp\left\{N \inf_{u<0} F^{-}_{M+1}(u)\right\}$$

where

$$\begin{cases} F^{+}_{M+1}(u) := \frac{1}{k+q}\psi_{M+1}\left(\frac{u}{P_0}\right) - (1+\alpha)\, u \\ F^{-}_{M+1}(u) := \frac{1}{k+q}\psi_{M+1}\left(\frac{u}{P_0}\right) - (1-\alpha)\, u \end{cases}$$

Once again we are interested in the Laplace transform of W_{M+1} which leads to study the behavior of the iterate of the function ψ.

4.3 The Laplace transform of W_{n+1}

Let $\beta := [k(1-P_0) + q(1-P_0 q)]/(k+q)$. We will proceed as explained in Section 3.2.1.

4.3.1 Estimates of ψ

Lemma 4.1 *Estimates of ψ for $u \geq 0$*
There exists $u_0 > 0$ such that

$$\psi(u) \leq \begin{cases} h\left(u; 1, \frac{\beta}{2}\right) \quad \textit{for} \;\; 0 \leq u \leq u_0 \\ (k+1)u \quad \textit{for} \;\; u \geq u_0 \end{cases}$$

Lemma 4.2 *Estimates of ψ for $u \leq 0$*
There exists $u^0 < 0$ such that

$$\psi(u) \leq h(u; 1, \beta) \quad \textit{for} \;\; u \leq u^0$$

4.3.2 Estimates of ψ_n

Proposition 9 *Estimates of ψ_n for $u \geq 0$*
Let $u_\star := h_n^{-1}(u_0; 1, \frac{\beta}{2})$.

$$\psi_n(u) \leq \begin{cases} h_n\left(u; 1, \frac{\beta}{2}\right) & \text{for } \ 0 \leq u \leq u_\star \\ u_0(k+1)^{n-\frac{2}{\beta}(\frac{1}{u}-\frac{1}{u_0})} & \text{for } \ u_\star \leq u \leq u_0 \\ (k+1)^n u \ \ \text{for } \ u_0 \leq u \end{cases}$$

Proposition 10 *Estimates of ψ_n for $u \leq 0$*

$$\psi_n(u) \leq h_n(u; 1, \beta) \ \ \text{for } \ u \leq u^0$$

4.4 Confidence interval

Finally, as in the deterministic case, we deduce

Theorem 4.1 *For α small enough, we have*

$$\mathbb{P}\left(\frac{|\hat{P} - \mathbb{P}(A)|}{\mathbb{P}(A)} \geq \alpha\right) \leq h_+ + h_- \tag{20}$$

where

$$\begin{cases} h_+ := \exp\left\{-\frac{2C}{(k+q)\beta[h(P_0)+P_0h(1)M/(M+1)]}\frac{1}{(M+1)^2}\left(\sqrt{1+\alpha}-1\right)^2\right\} \\ h_- := \exp\left\{-\frac{C}{(k+q)\beta[h(P_0)+P_0h(1)M/(M+1)]}\frac{1}{(M+1)^2}\left(1-\sqrt{1-\alpha}\right)^2\right\} \end{cases}$$

The first part of Proposition 3 follows from Theorem 4.1 using the same kind of argument as in Corollary 6 and provides better results than the results of the previous section since here we obtain an upper bound in $\exp\{\frac{1}{(M+1)^2}\}$ like in the optimal case.

5 Third model : sampling a random environment

5.1 The model

Here we decide to sample a random environment described by $(R_1, R_2, \ldots, R_M)$ at the beginning of the simulation, to keep the optimal values of P_i and M and to optimize the algorithm in $\mathbb{E}(R)$. More precisely, let $r_i = \{R_j^{(i)}, j = 1 \ldots M\}$, for $i = 1 \dots N$ N sequence of i.i.d. Bernoulli random variables $\{k, k+1\}$. r_i is said to be the i-th environmental sequence. Given that environmental sequences, we consider N independent branching processes $(Z_n^{(i)})_{n\geq 0}$, $i = 1, \dots, N$ in the same way as in the deterministic case except that each R_j is a random variable associated to a random generating function (g.f.) $f_j(s)$ which represents the offspring's distributions in j-th generation with $f_j(s) = (P_0 s + 1 - P_0)^{R_j}$.

(Z_n) is now a Branching Process in Randon Environment (BPRE).

As in usual case in random environments, we should be careful to distinguish the alea coming from the environment and the one from the process itself.

Since to each $(Z_n^{(i)})_{n\leq 0}$ $i = 1 \dots N$ of the $N > 1$ initial particles we associate a random environment r_i, the N initial particles reproduces independently one from each other. As a consequence, we can rewrite $\hat{P}$ the estimator of P as the sum of N independent branching processes :

$$\hat{P} = \frac{1}{N}\sum_{i=1}^{N} \tilde{P}_i = \frac{\mathbb{P}(A)}{N}\sum_{i=1}^{N} \frac{Z_{M+1}^{(i)}}{\xi_0\xi_1\dots\xi_M} = \frac{\mathbb{P}(A)}{N}\sum_{i=1}^{N} W_{M+1}^{(i)}$$

where $\xi_i = R_i P_0$ is a random variable for all $i = 1 \dots M$; and we aim at minimizing its variance for a fixed effort to derive the optimal parameter for R.

Some general results on BPRE

For more general background and details on BPRE, see for instance Athreya and Karlin [3, 4], Guivarc'h et al. [18], Smith and Wilkinson [27] and Tanny [28]. Suppose now that $r = \{r_j, j \geq 0\}$ is a general sequence of i.i.d. random variables that determines the succession of offspring g.f.'s $\{f_j(r_j; s), j \geq 0\}$,i.i.d., in a BPRE. Let us recall that, in complete analogy with the classical Galton-Watson process, a Galton-Watson process in random environment is subcritical if $\mathbb{E}(\log f_0'(1)) < 1$, critical if $\mathbb{E}(\log f_0'(1)) = 1$ and supercritical if $\mathbb{E}(\log f_0'(1)) > 1$. Moreover, in the subcritical and the critical (resp. supercritical) cases the probability of extinction given the environment r is one (resp. $q(r) \in [0, 1[$). Nevertheless in the critical case, under the assumption of a finite third moment for the random variables X_n, the asymptotic conditional distribution of $Z_n/\mathbb{E}(Z_n|Z_n > 0)$, given that $Z_n \neq 0$, is exponential, expressing the extreme character of this event for large n. In the supercritical case, W_n converges to some random variable W non-degenerated.

5.2 *The variance and its optimization*

5.2.1 *Study of the variance*

Proposition 11 *$\hat{P}$ is trivially an unbiased estimator and its variance is given by*

$$var(\hat{P}) = \frac{\mathbb{P}(A)^2}{N}\left(\frac{1}{P_0} - 1\right)\sum_{i=0}^{M}\mathbb{E}\left(\frac{1}{\xi}\right)^i \tag{21}$$

Proof First of all, $\operatorname{var}(\hat{P}) = \frac{\mathbb{P}(A)^2}{N}\operatorname{var}(W_{M+1})$. The calculation of $\operatorname{var}(\hat{P})$ then amounts to the calculation of $\operatorname{var}(W_{M+1})$.

Applying (17) to $X = W_{M+1} = \frac{Z_{M+1}}{P_0\prod_{i=1}^{M}\xi_i}$ and $\mathcal{F} = \sigma(Z_M, r)$ and since W_M is a martingale and $X_M^i \sim Bin(R_M, P_0)$, we get

$$\begin{aligned}
\operatorname{var}(W_{M+1}) &= \operatorname{var}(W_M) + \mathbb{E}\left(\frac{1}{[P_0\prod_{i=1}^{M}\xi_i]^2}\operatorname{var}(Z_{M+1}|\sigma(Z_M, r))\right)\\
&= \operatorname{var}(W_M) + \mathbb{E}\left(\frac{Z_M}{[P_0\prod_{i=1}^{M}\xi_i]^2}\operatorname{var}(X_M|\sigma(Z_M, r))\right)\\
&= \operatorname{var}(W_M) + \mathbb{E}\left(\frac{Z_M}{[P_0\prod_{i=1}^{M}\xi_i]^2}R_M P_0(1-P_0)\right)\\
&= \operatorname{var}(W_M) + (1-P_0)\mathbb{E}\left(\frac{W_M}{P_0\prod_{i=1}^{M}\xi_i}\right)
\end{aligned}$$

By a recurrent descent,

$$\operatorname{var}(W_{M+1}) = \operatorname{var}(W_0) + (1-P_0)\sum_{k=0}^{M}\mathbb{E}\left(\frac{W_k}{P_0\prod_{i=1}^{k}\xi_i}\right)$$

It remains to compute the expectation of $\frac{W_k}{P_0\prod_{i=1}^{k}\xi_i}$ which is derived by induction and at step k we have

$$\mathbb{E}\left(\frac{W_k}{P_0\prod_{i=1}^{k}\xi_i}\right) = \frac{1}{P_0}\mathbb{E}\left(\frac{1}{\xi}\right)^k \tag{22}$$

Clearly the formula holds for $k = 0 : \mathbb{E}\left(\frac{W_0}{P_0}\right) = \frac{1}{P_0}$. To go from $k-1$ to k, assume (22) for $k-1$.

$$\mathbb{E}\left(\frac{W_k}{P_0\prod_{i=1}^k \xi_i}\right) = \mathbb{E}\left(\frac{Z_{k-1}}{[P_0\prod_{i=1}^{k-1}\xi_i]^2\xi_k}\mathbb{E}(X_{k-1}|\sigma(Z_{k-1},r))\right)$$
$$= \mathbb{E}\left(\frac{W_{k-1}}{P_0\prod_{i=1}^k \xi_i}\right) = \mathbb{E}\left(\frac{W_{k-1}}{P_0\prod_{i=1}^{k-1}\xi_i}\right)\mathbb{E}\left(\frac{1}{\xi}\right)$$

since the R_i's are i.i.d. And we finally get the result using the induction's hypothesis.□

5.2.2 Optimization of the parameters

As done in Lagnoux [23], an optimal algorithm is chosen via the minimization of the variance of $\hat{P}$ for a given budget C, keeping the optimal values for M and P_0. The (average) cost is now

$$C = N\left[h(P_0)\sum_{i=0}^{M}\mathbb{E}(\xi)^i + Mh(1)\right] \tag{23}$$

Neglecting the cost introduced by the generation of the random environment, we assume, in the following,

$$C = N\left[h(P_0)\sum_{i=0}^{M}\mathbb{E}(\xi)^i\right], \qquad (h(1) \ll h(P_0)). \tag{24}$$

Once the trivial cases have been isolated, we can suppose $\mathbb{E}(R) \neq \frac{1}{P_0}$ and $\mathbb{E}(\frac{1}{R}) \neq P_0$, then the variance can be rewritten as

$$\mathrm{var}(\hat{P}) = \frac{\mathbb{P}(A)^2}{N}(\frac{1}{P_0} - 1)\frac{\mathbb{E}(\frac{1}{\xi})^{M+1} - 1}{\mathbb{E}(\frac{1}{\xi}) - 1}$$

and the cost has the following form

$$C = Nh(P_0)\frac{\mathbb{E}(\xi)^{M+1} - 1}{\mathbb{E}(\xi) - 1}$$

The optimal value of N is given by the cost

$$N = \frac{C}{h(P_0)}\frac{\mathbb{E}(\xi) - 1}{\mathbb{E}(\xi)^{M+1} - 1}$$

and we have to minimize the expression

$$\mathrm{var}(\hat{P}) = \frac{\mathbb{P}(A)^2}{C}h(P_0)\left(\frac{1}{P_0} - 1\right)\frac{\mathbb{E}(\frac{1}{\xi})^{M+1} - 1}{\mathbb{E}(\frac{1}{\xi}) - 1}\frac{\mathbb{E}(\xi)^{M+1} - 1}{\mathbb{E}(\xi) - 1}$$

under the constraint $\mathbb{E}(R) = 2k+1-k(k+1)\mathbb{E}(\frac{1}{R})$. To lead the analytic study, suppose that $1/P_0 = k + \delta$ with $\delta \in]0, 1[$ and let

– $u = \mathbb{E}(\xi) - 1$ and $v = \mathbb{E}(\frac{1}{\xi}) - 1$,
– $\rho = \frac{(k+\delta)^2}{k(k+1)}$ and $\alpha = \frac{\delta(1-\delta)}{k(k+1)}$.

The constraint becomes $v = \alpha - \rho u$ and by Lagrange multipliers, we finally need to solve

$$F(M, u) = \rho F(M, v) \tag{25}$$

where $F(M, u) = \frac{(1+u)^M (Mu-1)+1}{u[(1+u)^{M+1}-1]}$.

Remark 12 *Note that u and v depend on M and lie in the following intervals $[-\frac{\delta}{k+\delta}, \frac{1-\delta}{k+\delta}]$ and $[-\frac{1-\delta}{k+\delta}, \frac{\delta}{k}]$ respectively.*

We may only state asymptotic results.

Proposition 13 *Asymptotically in M, we only have three solutions :*
– $u \to u_1 = 0$ *and* $v \to v_1 = \alpha$
– $u \to u_2 = \frac{\alpha}{\rho}$ *and* $v \to v_2 = 0$
– $u \to u_3 = \frac{1+\alpha-\rho}{2\rho}$ *and* $v \to v_3 = \frac{\alpha+\rho-1}{2}$

Proof • Suppose first that u and v do not converge to 0 when $M \to +\infty$. So we have three cases to analyse (the case $u < 0, v < 0$ is not worth to consider since $\mathbb{E}(1/R)\mathbb{E}(R) \geq 1$) :
– $u < 0$ and $v > 0$
– $u > 0$ and $v < 0$
– $u > 0$ and $v > 0$

Case 1 : If $u < 0$ and $v > 0$ (or symetrically $u > 0$ and $v < 0$), since u does not converge to 0, there exists a subsequence of u such that $u < -\epsilon \; \forall M$. Then asymptotically in M, (25) is equivalent to

$$-\frac{1}{u} = \rho \frac{M}{1+v}$$

which is absurd since $v \neq -1$ and u does not converge to 0.

Case 2 : If $u > 0$ and $v > 0$, in the same way, there exists a subsequence of u and a subsequence of v such that $u > \epsilon$ and $v > \epsilon$. Then asymptotically in M, (25) is equivalent to

$$\frac{M}{1+u} = \rho \frac{M}{1+v}$$

And given the constraint, u must necessarily converge to $\frac{1+\alpha-\rho}{2\rho}$ and v to $\frac{\alpha+\rho-1}{2}$. Since the sequence u has only one adherence value and lies in a compact, it converges to $u_3 = \frac{1+\alpha-\rho}{2\rho}$ and similarly v converges to $v_3 = \frac{\alpha+\rho-1}{2}$. To guarantee $u > 0$ and $v > 0$, δ must be in $[\frac{k}{1+2k}, \frac{1}{2}] := [\delta_1, \delta_2]$.

• Suppose now that u converges to 0 when $M \to +\infty$. We have two cases to analyse : Mu bounded or not.

If Mu is not bounded, we can extract from Mu a subsequence which diverges. Then

$$F(M,u) = \frac{1 - \frac{1}{Mu} + \frac{1}{Mu(1+u)^M}}{1 - \frac{1}{(1+u)^{M+1}}} \frac{1}{u(1+u)} \sim \frac{1}{u(1+u)}$$

and thus (25) is equivalent to

$$\frac{1}{u(1+u)} = \rho \frac{M}{1+v}$$

which is absurd.

So Mu is bounded and we can extract of Mu a subsequence which converges to 0 or to a constant C_1. In the first case,

$$F(M,u) \sim \frac{e^{Mu}(Mu-1)+1}{u(e^{Mu}-1)} \sim \frac{(1+Mu)(Mu-1)+1}{u(1+Mu-1)} \sim M$$

and (25) is equivalent to

$$M = \rho \frac{M}{1+v}$$

which is absurd since $v > -1$. So the subsequence of Mu converges to a constant C_1 that solves

$$\frac{e^{C_1}(C_1-1)+1}{C_1(e^{C_1}-1)} = \frac{\rho}{1+\alpha}$$

Let $f(x) = \frac{e^x(x-1)+1}{x(e^x-1)} - \frac{\rho}{1+\alpha}$. Note that f is strictly increasing, converges to $1 - \frac{\rho}{1+\alpha} \geq 0$ when $x \to +\infty$ and to $-\frac{\rho}{1+\alpha} \leq 0$ when $x \to -\infty$ for $\delta \leq \delta_2$. So C_1 is defined uniquely and then the sequence Mu has only one limiting point and lies in a compact, so it converges to C_1, u converges to $u_1 = 0$ and v converges to $v_1 = \alpha$.

Symetrically we get the equivalent result for v : it converges to $v_2 = 0$ and u converges to $u_2 = \frac{\alpha}{\rho}$. $\square$

Remind that we want to determine the minimum of the variance of $\hat{P}$.

Proposition 14 *Let $\delta_0 := \sqrt{k}(\sqrt{k+1} - \sqrt{k})$. Then asymptotically in M,*
– *if $\delta \in [0, \delta_0]$, (u_1, v_1) minimizes the variance,*
– *if $\delta \in [\delta_0, 1]$, (u_2, v_2) minimizes the variance.*

Proof Let $\beta = \frac{P^2}{C} h(P_0) \left(\frac{1}{P_0} - 1\right)$ and for $i = 1, 2, 3$,

$$\text{var}_i = \beta \frac{[1+u_i]^{M+1} - 1}{u_i} \frac{[1+v_i]^{M+1} - 1}{v_i}$$

First of all, compare var_1 and var_2. For $i = 1, 2$,

$$\text{var}_i \sim \beta \frac{e^{(M+1)(1+C_i-1/M)} - 1}{u_i} \frac{[1+v_i]^{M+1}}{v_i} \sim \beta M \frac{e^{C_i} - 1}{C_i} \frac{[1+v_i]^{M+1}}{v_i}$$

Thus

$$\frac{\text{var}_1}{\text{var}_2} \sim \frac{e^{C_1} - 1}{C_1} \frac{C_2}{e^{C_2} - 1} \frac{u_2}{v_1} \left(\frac{1+v_1}{1+u_2}\right)^{M+1}$$

But $\frac{1+v_1}{1+u_2} < 1 \Leftrightarrow \delta \le \delta_0 := \sqrt{k}(\sqrt{k+1} - \sqrt{k})$. As a conclusion,

$$\text{var}_1 \le \text{var}_2 \Leftrightarrow \delta \le \delta_0$$

Then note that var_3 is not defined on $[0, \delta_1] \cup [\delta_2, 1]$ and δ_0 trivially belongs to $]\delta_1, \delta_2[$. Consequently, for $\delta \in [\delta_1, \delta_2]$, it is sufficient to compare
– var_1 and var_3 on $[\delta_1, \delta_0]$,
– var_2 and var_3 on $[\delta_0, \delta_2]$.

But $\frac{\text{var}_1}{\text{var}_3} \sim M \frac{e^C - 1}{C} \frac{u_3 v_3}{\alpha} \left[4\rho \frac{1+\alpha}{(1+\alpha+\rho)^2} \xrightarrow[M\to\infty]{}\right]^{M+1}$ and $4\rho \frac{1+\alpha}{(1+\alpha+\rho)^2} = \frac{4\frac{\rho}{1+\alpha}}{(1+\frac{\rho}{1+\alpha})^2} < 1$. Thus $\frac{\text{var}_1}{\text{var}_3} \xrightarrow[M\to\infty]{} 0$ and $\text{var}_1 \le \text{var}_3$ asymptotically in M. In the same way, $\frac{\text{var}_2}{\text{var}_3} \xrightarrow[M\to\infty]{} 0$. □

5.2.3 Confidence interval

By analogy with the deterministic case, we are interested in bounds on

$$\mathbb{P}\left(|\hat{P} - \mathbb{P}(A)|/\mathbb{P}(A) \ge \alpha\right)$$

So let $\psi_i(u) := R_i \log(P_0 e^u + 1 - P_0)$, $f_{0,i} := f_0 \circ \ldots \circ f_i$ and $\psi_{0,i} := \psi_0 \circ \ldots \circ \psi_i$, taking again $\xi_i = R_i P_0$ for all $i = 1, \ldots, M$ and $\xi_0 = P_0$,

$$\mathbb{E}\left(e^{uW_{M+1}}\right)=\mathbb{E}\left[P_0 f_{1,M}\left(e^{\frac{u}{\prod_{i=0}^{M}\xi_i}}\right)+1-P_0\right]=\mathbb{E}\left[\exp\left\{\psi_{0,M}\left(u/\prod_{i=0}^{M}\xi_i\right)\right\}\right]$$

Finally,

$$\mathbb{P}\left(\frac{|\hat{P}-\mathbb{P}(A)|}{\mathbb{P}(A)}\geq\alpha\right)\leq\mathbb{E}\left(\exp\left\{\inf_{u>0}F^{+}_{M+1}(u)\right\}\right)^{N}+\mathbb{E}\left(\exp\left\{\inf_{u<0}F^{-}_{M+1}(u)\right\}\right)^{N}$$

where

$$\begin{cases} F^{+}_{M+1}(u):=\psi_{0,M}\left(\frac{u}{\prod_{i=0}^{M}\xi_i}\right)-(1+\alpha)\,u \\ F^{-}_{M+1}(u):=\psi_{0,M}\left(\frac{u}{\prod_{i=0}^{M}\xi_i}\right)-(1-\alpha)\,u \end{cases}$$

And we are interested in the Laplace transform of W_{M+1} which leads to study the behavior of the iterate of the random functions ψ.

5.3 The Laplace transform of W_{n+1}

5.3.1 Criticality

By simple arguments of convexity,

Proposition 15 *Asymptotically in M,*
– for all $\delta\in[0,\delta_0]$, we are in the subcritical case,
– for all $\delta\in[\delta_0,1]$, we are in the supercritical case.

Proposition 16 *The critical case is given by $p_c:=\mathbb{P}(R=k)$*

$$p_c=\frac{\log\left(\frac{k+\delta}{k+1}\right)}{\log\left(\frac{k}{k+1}\right)}$$

and - $p>p_c\Leftrightarrow\mathbb{E}(\log f_1'(1))<0\Leftrightarrow$ subcritical case,
- $p=p_c\Leftrightarrow\mathbb{E}(\log f_1'(1))=0\Leftrightarrow$ critical case,
- $p<p_c\Leftrightarrow\mathbb{E}(\log f_1'(1))>0\Leftrightarrow$ supercritical case.

Note that

$$p\underset{M\to\infty}{\to}p_a:=\begin{cases}1-\delta & \text{for }\ \delta<\delta_0\\ \frac{k(1-\delta)}{k+\delta} & \text{for }\ \delta>\delta_0\end{cases}$$

Let us plot p_a and p_c in Figure 4 to obtain in a different way the distinction between the supercritical case and the subcritical one.

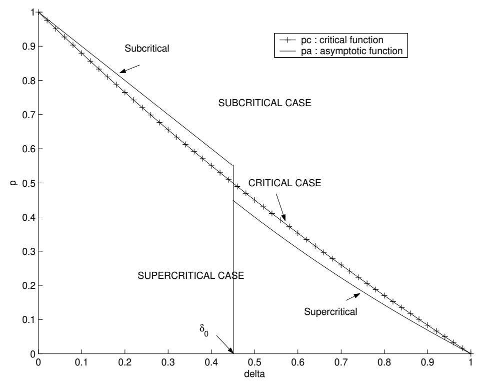


FIG. 4. Plot p=f(δ) asymptotically in M

5.3.2 Estimates of $\psi_{0,M}$

Remind we aim at estimate $\psi_{0,M} = \psi_0 \circ \ldots \circ \psi_M$ and so we have to iterate

$$\begin{cases} \psi^{(1)}(u) = k \log(P_0 e^u + 1 - P_0) \\ \quad \text{and} \\ \psi^{(2)}(u) = (k+1) \log(P_0 e^u + 1 - P_0) \end{cases}$$

Heuristic : Whatever the case studied ($m > 1$, $m < 1$), we simply use two bounding functions for ψ. More precisely, for example, for $u \geq 0$, until some a (intersection between the lowest homography ($m < 1$) and the first bissector) we bound ψ by the homographic function and then by the asymptotic direction in both cases.

• For $u \geq 0$. Let
– $m^{(1)} = kP_0 < 1$ and $m^{(2)} = (k+1)P_0 > 1$,
– $a^+ := \arg_{u \neq 0}\{h(u; m^{(1)}, \frac{1-P_0}{2}) = u\} = 2\frac{1-m^{(1)}}{1-P_0}$,
– $b^+ := h(a^+; m^{(2)}, \frac{1-P_0}{2})$,
– $u^+ := h^{-1}_{M+1}(a^+; m^{(2)}, \frac{1-P_0}{2})$.

Following Section 3.2.1 in a simplified version, we deduce

$$\psi^{(1)}(u) \leq \begin{cases} h\left(u; m^{(1)}, \frac{1-P_0}{2}\right) & \text{for } \ 0 \leq u \leq a^+ \\ g(u; k, a^+(1-k)) & \text{for } \ u \geq a^+ \end{cases}$$

and

$$\psi^{(2)}(u) \leq \begin{cases} h\left(u; m^{(2)}, \frac{1-P_0}{2}\right) & \text{for } \; 0 \leq u \leq a^+ \\ g(u; k+1, b^+ - a^+(k+1)) & \text{for } \; u \geq a^+ \end{cases}$$

which leads to the following proposition

Proposition 17 *If $a_0 = a^+$ and for all $i = 1 \ldots M$,*

$$a_i := \begin{cases} a^+ & \textit{if} \;\; R_i = k \\ b^+ & \textit{if} \;\; R_i = k+1 \end{cases}$$

$$\psi_{0,M}(u) \leq \begin{cases} h^+_{0,M}(u) & \textit{for} \;\; 0 \leq u \leq u^+ \\ g^+_M(u) & \textit{for} \;\; u^+ \leq u \leq a^+ \\ \prod_{i=0}^M R_i u + \sum_{i=0}^M \left(\prod_{j=0}^{i-1} R_j\right) a_i - \sum_{i=0}^M \left(\prod_{j=0}^{i} R_j\right) a^+ & \textit{for} \;\; u \geq a^+ \end{cases}$$

where
– $h^+_{0,M}(u) = u\left(\prod_{i=0}^M \xi_i\right) / \left[1 - \frac{1-P_0}{2} u \left(1 + \sum_{i=1}^M \xi_i \ldots \xi_M\right)\right]$ *and*
– $g^+_M(u) = \begin{cases} h^+_{0,M}(u) & \textit{w.p.} \;\; \mathbb{P}\left(0 \leq h^+_{0,M}(u) \leq a^+\right) \\ b^+ & \textit{w.p.} \;\; 1 - \mathbb{P}\left(0 \leq h_{0,M}(u) \leq a^+\right) \end{cases}$

• For $u \leq 0$. Let
– $a^- := \arg_{u \neq 0}\{h(u; m^{(2)}, 1-P_0) = u\} = \frac{1-m^{(2)}}{1-P_0}$,
– $b^- := h(a^-; m^{(1)}, 1-P_0)$,
– $u^- := h^{-1}_{M+1}(a^-; m^{(1)}, 1-P_0)$.

In the same way, we deduce

$$\psi^{(1)}(u) \leq \begin{cases} h\left(u; m^{(1)}, 1-P_0\right) = \frac{m^{(1)}u}{1-(1-P_0)u} & \text{for } \; a^- \leq u \leq 0 \\ b^- \;\; \text{for } \; u \leq a^- \end{cases}$$

and

$$\psi^{(2)}(u) \leq \begin{cases} h\left(u; m^{(2)}, 1-P_0\right) & \text{for } \; a^- \leq u \leq 0 \\ b^- \;\; \text{for } \; u \leq a^- \end{cases}$$

which leads to the following proposition

Proposition 18

$$\psi_{0,M}(u) \leq \begin{cases} h^-_{0,M}(u) & \textit{for} \;\; u^- \leq u \leq 0 \\ g^-_M(u) & \textit{for} \;\; a^- \leq u \leq u^- \\ b^- & \textit{for} \;\; u \leq a^- \end{cases}$$

where

– $h^-_{0,M}(u) = u\left(\prod_{i=0}^M \xi_i\right) / \left[1-(1-P_0)u\left(1+\sum_{i=1}^M \xi_i\ldots\xi_M\right)\right]$ *and*

– $g^-_M(u) = \begin{cases} h^-_{0,M}(u) & w.p.\ \ \mathbb{P}\left(b \le h^-_{0,M}(u) \le 0\right) \\ b^- \ \ w.p.\ \ 1-\mathbb{P}\left(b \le h^-_{0,M}(u) \le 0\right) \end{cases}$

5.3.3 About random walk on the affine group and consequences

We would like to estimate

$$\mathbb{P}\left(0 \le h^+_{0,M}(u) \le a\right) \ \text{ for } \ u \ge 0 \ \text{ and } \ \mathbb{P}\left(b \le h^-_{0,M}(u) \le 0\right) \ \text{ for } \ u \le 0$$

Remind $\xi_i = R_i P_0$, let $v = \frac{1-P_0}{2}$ and

$$\begin{cases} x_{n+1}(y_0) = \xi_0\ldots\xi_n y_0 + v[1+\sum_{i=1}^n \xi_i\ldots\xi_n] \\ y_{n+1}(y_0) = \xi_0\ldots\xi_n y_0 + v[1+\sum_{i=0}^{n-1} \xi_0\ldots\xi_i] \end{cases}$$

Random walk on the affine group : Consider the affine transformations $g_k(x) = \xi_k x + v$ ($x \in \mathbb{R}$) and the random walk $(g_0 \circ g_1 \circ \ldots \circ g_n)$ on the affine group. Immediatetly

$$\begin{cases} g_0 \circ g_2 \circ \ldots \circ o g_n(x) = y_{n+1}(x) \\ g_n \circ g_{n-1} \circ \ldots \circ g_0(x) = x_{n+1}(x) \end{cases}$$

and

Lemma 5.1 *(i)* $y_n \stackrel{\mathcal{L}}{=} x_n$
(ii) $\{0 \le h_{0,n}(u) \le a\} = \{0 \le y_{n+1}(1/a) \le 1/u\}$
(iii) $\{b \le h_{0,n}(u) \le 0\} = \{\frac{1}{u} \le y_{n+1}(1/b) \le 0\}$

Hence the link between the random walk on the group of the homographic transformations and the one on the affine group.

Therefore, we are interested in the asymptotic properties of $g_0 \circ g_1 \circ \ldots \circ g_n$'s distribution which is obviously characterized by the law of ξ. As for the BPRE, we distinguish three different regimes determined by the position of $\mathbb{E}(\log \xi)$ with respect to the origin.

For general details on random walk on affine groups the reader is referred for instance to Vervaat [29], Brandt [9] for the one-dimensional case and to Kesten [22], Bougerol and Picard [8], Babillot et al. [6] for the d-dimensional case.

Subcritical case : $\mathbb{E}(\log \xi) < \mathbf{0}$
First of all, Diaconis and Freedman [12] allow us to state that the backward process $\psi_{0,M}(u) = \psi_0 \circ \ldots \circ \psi_M(u)$ converges almost surely, at an exponential rate to a random limit that does not depend on the starting point u.

Concerning the random walk on the affine group in our particular case ($\xi \stackrel{\mathcal{L}}{=} Ber$ on $\{kP_0, (k+1)P_0\}$), we have by Maksimov [25]

$$y_n \xrightarrow[n\to\infty]{} v(1+D)$$

where $D = \sum_{i=0}^{\infty} \xi_0 \ldots \xi_i$. In general, nothing is known on the distribution of D except that it satisfies the following functional equation

$$F(\lambda) = pF\left(\frac{\lambda}{kP_0} - 1\right) + (1-p)F\left(\frac{\lambda}{(k+1)P_0} - 1\right)$$

with $p = \mathbb{P}(R = k)$.

But Kesten [22] gives an asymptotic result on the distribution of D : for some $C > 0$ and some $\kappa > 0$,

$$\mathbb{P}(D \geq x) \underset{x\to\infty}{\sim} \frac{C}{x^\kappa}$$

Corollary 19 *Letting* $\kappa = -\log(p(1-p))/\log(k(k+1)P_0^2)$,

$$\mathbb{P}\left(0 \leq h_{0,n}(u) \leq a\right) \underset{n\to\infty}{\sim} \mathbb{P}\left(v(1+D) \leq \frac{1}{u}\right) \underset{n\to\infty, u\to 0}{\sim} 1 - \frac{C}{(\frac{1}{uv} - 1)^\kappa}$$

Critical case : $\mathbb{E}(\log \xi) = \mathbf{0}$
Maksimov [25] asserts that the distribution of $\log y_n/\sqrt{n}$ approaches, as $n \to \infty$, the truncated normal distribution for any starting point of the walk.

Corollary 20 *Let* Z *a truncated normal distribution and* $\sigma^2 = var(\xi)$.

$$\mathbb{P}\left(0 \leq h_{0,n}(u) \leq a\right) \underset{n\to\infty}{\sim} \mathbb{P}\left(Z \leq -\frac{1}{\sqrt{n}} \log u\right) \underset{n\to\infty}{\sim} \left[1 - e^{-\frac{\log u^2}{n}}\right]^{\frac{1}{2}}$$

Supercritical case : $\mathbb{E}(\log \xi) > \mathbf{0}$
Following Maksimov [25], the distribution of $(\frac{y_n}{v})^{1/\sqrt{n}} e^{-M\sqrt{n}}$ approaches, as n goes ∞, the log normal distribution for any starting point of the walk.

Corollary 21 *Let* $Z \stackrel{\mathcal{L}}{=} N(0,1)$ *and* $\sigma^2 = var(\xi)$.

$$\mathbb{P}\left(0 \leq h_{0,n}(u) \leq a\right) \underset{n\to\infty}{\sim} \frac{1}{2}\exp\left\{-\frac{1}{2\sigma^2}n\left[\frac{1}{n}\log(uv) + \mathbb{E}(\log\xi)\right]\right\}$$

5.4 Confidence interval

Finally, as in the deterministic case, we deduce

Theorem 5.1 *For* α *small enough,*

$$\mathbb{P}\left(\frac{|\hat{P} - \mathbb{P}(A)|}{\mathbb{P}(A)} \geq \alpha\right) \leq h_+^N + h_-^N \tag{26}$$

where

$$\begin{cases} h_+ := \mathbb{E}\left(\exp\left\{-\frac{2}{1-P_0}\frac{1}{\sum_{i=0}^M \xi_0^{-1}\ldots\xi_i^{-1}}\left(\sqrt{1+\alpha}-1\right)^2\right\}\right) \\ h_- := \mathbb{E}\left(\exp\left\{-\frac{1}{1-P_0}\frac{1}{\sum_{i=0}^M \xi_0^{-1}\ldots\xi_i^{-1}}\left(1-\sqrt{1-\alpha}\right)^2\right\}\right) \end{cases}$$

where $N = \frac{C}{h(P_0)}\frac{\mathbb{E}(\xi)-1}{\mathbb{E}(\xi)^{M+1}-1}$. *By Jensen's inequalities,*

$$\mathbb{P}\left(\frac{|\hat{P} - \mathbb{P}(A)|}{\mathbb{P}(A)} \geq \alpha\right) \leq g_+^N + g_-^N \tag{27}$$

where

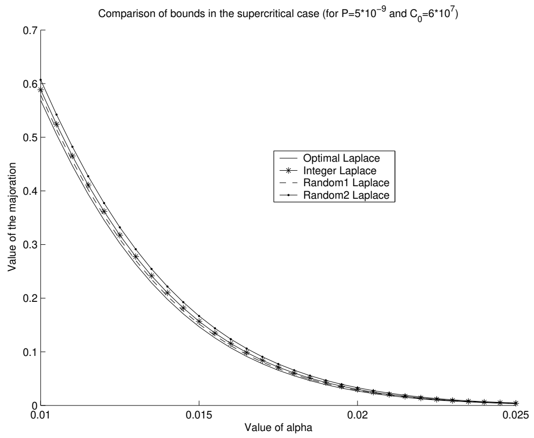
$$\begin{cases} g_+ := \exp\left\{-\frac{2}{1-P_0}\frac{1}{\mathbb{E}(\xi^{-1})}\frac{\mathbb{E}(\xi^{-1})-1}{\mathbb{E}(\xi^{-1})^{M+1}-1}\left(\sqrt{1+\alpha}-1\right)^2\right\} \\ g_- := \exp\left\{-\frac{1}{1-P_0}\frac{1}{\mathbb{E}(\xi^{-1})}\frac{\mathbb{E}(\xi^{-1})-1}{\mathbb{E}(\xi^{-1})^{M+1}-1}\left(1-\sqrt{1-\alpha}\right)^2\right\} \end{cases}$$

Using the same kind of argument as in Corollary 6, we obtain the second part of Proposition 3.

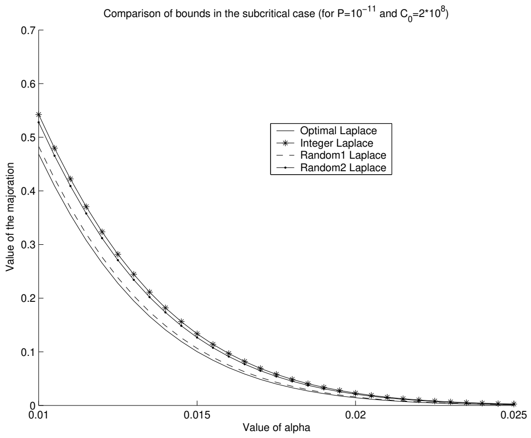
6 Numerical illustration

We plot in Figure 5 the bounds given by the Laplace transform in the different models :
– in the optimal model : $m = 1$ (plain line),
– in the deterministic model where $m \neq 1$ ($- * -$ line),
– in the model where we sample a new R at each success (dashed line),
– in the model where we sample a random environment ($-\cdot-$ line),

in two different cases
– for $\mathbb{P}(A) = 5\ 10^{-9}$ and $C = 6\ 10^7$, then m=1.055 (supercritical case),
– for $\mathbb{P}(A) = 10^{-11}$ and $C = 2\ 10^8$, then m=0.924 (subcritical case).



1. Supercritical case



2. Subcritical case

FIG. 5. Confidence interval

7 Conclusion

In this article, the relative simplicity of our model allows us to state explicit results as the Chernoff's bounds of the relative error between the estimate $\widehat{P}$ and $\mathbb{P}(A)$. Going further in the calculus, using the heuristic presented above, one can also deduce a central limit theorem and so Berry-Esseen bounds. Then we study the sensitivity of the Chernoff's bounds depending on the choice of the splitting number R in three different algorithms based on the branching splitting model : when one can not be exactly in the critical case (which corresponds to the optimal algorithm), the best way to proceed is to consider R as a random variable that we generate at each success during the simulation. Besides, this procedure is currently used in practice, see for example [2].

In practice, we do not know the transition probablities but just empirical estimation on them, and we can bound to adjust the levels according to them. In a model where the P_i are unknown but belong in some known interval, we may proceed in the following way :

(1) Choose an arbitrarly sequence $\left(R_1^{(0)}, R_2^{(0)}, \dots R_M^{(0)}\right)$ of sampling numbers. During the first step, sample a packet of $\theta_{1,N}N$ particles following the splitting algorithm with sampling numbers $R_i^{(0)}$. Thus empirical estimations $\left(\hat{P}_i^{(1)}\right)_{i=1\dots M+1}$ of $(P_i)_{i=1\dots M+1}$ are derived.

(2) Compute the new sampling numbers $\left(R_i^{(1)}\right)_{i=1\ldots M}$ as suggested in the algorithm optimization

$$R_i^{(1)} = \frac{1}{\sqrt{\hat{P}_i^{(1)}\hat{P}_{i+1}^{(1)}}}\sqrt{\frac{1-\hat{P}_{i+1}^{(1)}}{1-\hat{P}_i^{(1)}}}$$

During the second step, sample a second packet of $theta_{2,N}N$ particles following the splitting algorithm with sampling numbers $R_i^{(1)}$. Thus empirical estimations $\left(\hat{P}_i^{(2)}\right)_{i=1\ldots M+1}$ (better than the first ones) of $(P_i)_{i=1\ldots M+1}$ are derived.

(3) Repeat that procedure until the budget is entirely consumed.

The goal of this algorithm is to be as close as possible to the optimal algorithm. The precise study of the proportion of the budget to use in each step shall be the purpose of a forthcoming paper and is not derive straightforward.

Acknowledgement

I gratefully acknowledge my supervisors Dominique Bakry and Pascal Lezaud for their helpful suggestions and comments.

Références

[1] Aldous, David. Probability Approximations via the Poisson Clumping Heuristic, volume 77 of *Applied Mathematical Sciences*. Springer Verlag (1989).

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

[3] K.B.Athreya and S. Karlin, On branching processes with random environments. I. Extinction probabilities, Ann. Math. Statist, 42 (1971) 1499-1520.

[4] K.B.Athreya and S. Karlin, Branching processes with random environments. II. Limit theorems, Ann. Math. Statist., 42 (1971) 1843-1858.

[5] K.B. Athreya and P.E. Ney, Branching processes, Die Grundlehren der mathematischen Wissenschaften, Band 196, Springer-Verlag, 1972.

[6] M. Babillot and P. Bougerol and L. Elie, The random difference equation $X_n = A_nX_{n-1} + B_n$ in the critical case, Ann. Probab., 25 (1997) 1 478-493.

[7] P.T. de Boer, Analysis and efficient simulation of queueing models of telecommunication systems, PhD thesis, University of Twente, The Nederlands, 2000.

[8] P. Bougerol and N. Picard, Strict stationarity of generalized autoregressive processes, Ann. Probab., 20 (1992) 4 1714-1730.

[9] A. Brandt, The stochastic equation $Y_{n+1} = A_n Y_n + B_n$ with stationary coefficients, Adv. in Appl. Probab., 18 (1986)1 211-220.

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

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

[12] P. Diaconis and D. Freedman, Iterated random functions, SIAM Review, 41 (1999) 1 45-76.

[13] P. Diaconis and S. Holmes, Three examples of Monte-Carlo Markov chains : at the interface between statistical computing, computer science, and statistical mechanics, Discrete probability and algorithms (Minneapolis, MN, 1993), IMA Vol. Math. Appl., 72 (1995) 43-56.

[14] A. Doucet and N. de Freitas and N. Gordon, An introduction to sequential Monte Carlo methods, Sequential Monte Carlo methods in practice, Stat. Eng. Inf. Sci., Springer, (2001) 3-14.

[15] M.J.J. Garvels, The splitting method in rare event simulation, PhD thesis, University of Twente, The Nederlands, 2000.

[16] P. Glasserman and P. Heidelberger and P. Shahabuddin and T. Zaji, A large deviations perspective on the efficiency of multilevel splitting, IEEE Trans. Automat. Control, 43 (1998)12 1666-1679.

[17] P. Glasserman and P. Heidelberger and P. Shahabuddin and T. Zajic, Multilevel splitting for estimating rare event probabilities, Operations Research, 47 (1999) 4 585-600.

[18] Y. Guivarc'h and E. Le Page and Q. Liu, Normalisation d'un processus de branchement critique dans un environnement aléatoire, C. R. Math. Acad. Sci. Paris, 337 (2003) 9 603-608.

[19] T.E. Harris, The theory of branching processes, Dover Phoenix Editions, Corrected reprint of the 1963 original [Springer, Berlin ; MR**29** #664], 2002.

[20] P. Heidelberger, Fast Simulation of Rare Events in Queueing and Reliability Models, ACM Transactions on Modeling and Simulation, 5(1) (1995) 43-85.

[21] M. Jerrum and A. Sinclair, The Markov chain Monte Carlo method : an approach to approximate counting and integration, Approximation Algorithms for NP-hard Problems, (1997) 482-520.

[22] H. Kesten, Random difference equations and renewal theory for products of random matrices, Acta Mathematica Academiae Scientiarum Hungaricae, 131 (1973) 207-248.

[23] A. Lagnoux, Rare event simulation, Probability in the Engineering and Informational Sciences, 20 (2006) 45-66.

[24] R. Lyons, Probability on trees and networks. *Preprint.* http ://mypage.iu.edu/ rdlyons/, 2002.

[25] V.M. Maksimov, A generalized Bernoulli scheme and its limit distributions, Teor. Verojatnost. i Primenen., 18 (1973) 547-556.

[26] J.S. Sadowsky, On Monte Carlo estimation of large deviations probabilities, Ann. Appl. Probab., 6 (1996) 2 399-422.

[27] W.L. Smith and W.E. Wilkinson, On branching processes in random environments, Ann. Math. Statist., 40 (1969) 814-827.

[28] D. Tanny, Limit theorems for branching processes in a random environment, Ann. Probability, 5 (1977) 1 100-116.

[29] W. Vervaat, On a stochastic difference equation and a representation of nonnegative infinitely divisible random variables, Advances in Applied Probability, 11 (1979) 4 750-783.

[30] M. Villén-Altamirano and J. Villén-Altamirano, RESTART : a Method for Accelerating Rare Event Simulations, 13th International Teletraffic Congress, Copenhagen, (1991) 71-76.

[31] Villén-Altamirano, Manuel and Villén-Altamirano, José, Restart : An Efficient and General Method for Fast Simulation of Rare Event, Tech. Rept. No 7, Departamento de Mathematica Aplicada, E.U. Informática, Universidad Polytéchnica de Madrid, 1997.

List of figures

List of tables