Multilevel branching splitting algorithm for estimating rare event probabilities

Agnès Lagnoux, Pascal Lezaud

Multilevel branching and splitting algorithm for estimating rare event probabilities

Agnès Lagnoux[a], Pascal Lezaud[b]

[a]*Institut de Mathématiques de Toulouse, Université Toulouse 2 Jean Jaurès, France*

[b]*ENAC - Ecole Nationale de l'Aviation Civile, Université de Toulouse, France*

Abstract

We analyse the splitting algorithm performance in the estimation of rare event probabilities in a discrete multidimensional framework. For this we assume that each threshold is partitioned into disjoint subsets and the probability for a particle to reach the next threshold will depend on the starting subset. A straightforward estimator of the rare event probability is given by the proportion of simulated particles for which the rare event occurs. The variance of this estimator is the sum of two parts: with one part resuming the variability due to each threshold, and the second part resuming the variability due to the number of thresholds. This decomposition is analogous to that of the continuous case. The optimal algorithm is then derived by cancelling the first term leading to optimal thresholds. Then we compare this variance with that of the algorithm in which one of the threshold has been deleted. Finally, we investigate the sensitivity of the variance of the estimator with respect to a shape deformation of an optimal threshold. As an example, we consider a two-dimensional Ornstein-Uhlenbeck process with conformal maps for shape deformation.

Keywords: splitting, rare event probability estimation, Monte Carlo, branching process, simulation, variance reduction, first crossing time density, conformal maps

1. Introduction

The risk modelling approach consists of, first, formalizing the system under consideration, and second, using mathematical or simulation tools to obtain some estimates (Aldous, 1989; Sadowsky, 1996). Analytical and numerical approaches are useful, but may require many simplifying assumptions. On the other hand, a Monte Carlo simulation is a practical alternative when the analysis requires fewer simplifying assumptions. Nevertheless, obtaining accurate estimates of rare event probabilities, say about 10^{-9} to 10^{-12}, using traditional techniques requires an extremely large amount of computing time.

Many techniques for reducing the number of trials in a Monte Carlo simulation have been proposed, such as importance sampling or trajectory splitting (L'Ecuyer et al. (2009)). In the splitting technique, we suppose that there exists some well-identifiable intermediate states that are visited more frequently than the target states themselves, and behave as gateways to reach the rare event. Thus, we consider a decreasing sequence of events B_i leading to the rare event B:

$$B \coloneqq B_{M+1} \subset B_M \subset \ldots \subset B_1 \,. \tag{1}$$

Then $p \coloneqq \mathbb{P}(B) = \mathbb{P}(B|B_M)\mathbb{P}(B_M|B_{M-1}) \ldots \mathbb{P}(B_2|B_1)\mathbb{P}(B_1)$, where on the right-hand side, each conditioning event is “not rare”. These conditional probabilities are, in general, not available explicitly. Instead, we know how to make the particles evolve from level B_i to the next level B_{i+1} (e.g. Markovian behaviour).

The principle of the algorithm is first to simultaneously run several particles starting from level B_i. After a period of time, some have evolved "badly", while others have evolved "well", that is, they have succeeded in reaching the threshold B_{i+1}. Then the "bad" particles are moved to the position of the "good" ones, and so on, until B is reached. In such a way, the more promising particles are favoured. Examples of this class of algorithms can be found in Aldous and Vazirani (1994), with their "go with the winners" scheme, in Jerrum and Sinclair (1997) and Diaconis and Holmes (1995) with approximate counting, and in a more general setting in Doucet et al. (2001); Del Moral (2004); Cérou and Guyader (2005); Del Moral and Garnier (2005); Morio et al. (2014).

The difficulty comes from the complexity of the dynamics of the particles. A simpler analysis can be completed by focusing only on the underlying Markov chain that represents the changes of thresholds. In this technique, we make a Bernoulli trial to check whether or not the set event B_1 has occurred. In that case, we split this trial into R_1 Bernoulli subtrials, and for each of them we again check whether or not the event B_2 has occurred. This procedure is repeated at each level, until B is reached. If an event level is not reached, then neither is B, and we stop the current retrial. Using N independent replications of this procedure, we have 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 potential $R_i \dots R_M$ retrials have failed. Clearly the particles reproduce and evolve independently.

An unbiased estimator of p is given by the quantity:

$$\widehat{p}_{M+1} = \frac{N_B}{N \prod_{i=1}^{M} R_i},$$

where N_B is the total number of trajectories having reached the set B. Considering that this algorithm is represented by N independent Galton-Watson branching processes, as seen in Lagnoux (2006), the variance of $\widehat{p}_{M+1}$ can then be derived, and depends on the probability transitions and the mean numbers of particles' successes at each level. Leading by the heuristic presented in Villén-Altamirano and Villén-Altamirano (1991, 1997), an optimal algorithm is derived by minimising the variance of the estimator for a given budget (or computational cost). This cost is defined as the expected number of trials generated during the simulation, with each trial weighted by a cost function.

The optimisation of the algorithm suggests that all transition probabilities be taken as equal to a constant, and the numbers of splitting be equal to the inverse of this constant Lagnoux (2006). Then we deduce the number of thresholds M, and finally the number N of replications. In fact, optimal values are chosen in such a way so as to balance between the increase of the variance when the number splitting is small and the exponential growth in computational effort when too much splitting is used.

In the higher dimension, the engineering community has proposed algorithms to estimate rare event probabilities. Subset simulation, which is also based on a partitioning of the space into nested subsets, uses the Markov Chain simulation (in particular, the Metropolis Hastings scheme), see Au and Beck (2001). Importance sampling techniques have also been developed in that framework. When the failure region is not too complex to describe, schemes to construct importance sampling algorithms have been introduced that are based on design points (see e.g., Au et al. (1999), and Kiureghian and Dakessian (1998), and the references therein), or adaptive pre-samples (see e.g., Au and Beck (1999) and the references therein). When the complexity of the rare event increases, it appears to be difficult to construct an efficient importance sampling scheme (Schuëller et al. (1993)).

In this paper, we continue the multidimensional approach and theoretically study the algorithm introduced in Glasserman et al. (1998) and Garvels (2000), particularly in order to obtain a new expression of the variance of the estimator analogous to that of the continuous case (L'Ecuyer et al. (2009)). Thus, we assume that each threshold is partitioned into s disjoint subsets and the probability for a particle starting from a threshold to reach the next threshold will depend on the starting subset. Unlike the unidimensional case, the difficulty of reaching the next threshold differs according to the starting subset; it may be that the threshold is no longer an iso-probability level. In this context, the variance of the estimator $\widehat{p}_{M+1}$ is the sum of two parts: with one part resuming the variability due to each threshold and the second part resuming the variability due to the thresholds' number (see Proposition 3.1). For the unidimensional case, only the second term remains. The optimal algorithm is then derived by cancelling the first term of the variance leading to iso-probability levels and by optimising the other parameters as in the unidimensional case.

Furthermore, by introducing new operators, we obtain an alternative expression of the variance which is more tractable when we wish to compare the variance of the estimators in an algorithm with M thresholds with the variance in an algorithm in which one of the threshold has been deleted. More precisely, we study the need for an intermediate threshold and derive a procedure to detect whether or not to keep it. In order to obtain a simple criteria, we assume the optimal shape of the thresholds of the optimal algorithm. Finally, we investigate the sensitivity of the variance of $\widehat{p}_{M+1}$ with respect to a shape deformation of the threshold, relative to the optimal shape.

The remainder of this paper is divided into five sections. In Sections 2–4, we present, theoretically analyse and optimise the splitting algorithm in the multidimensional case. Next, Sections 5 and 6 deal with the sensitivity analysis of the variance as previously presented. In particular, in Section 6, we illustrate a way by which to deform the shape of the thresholds in order to get uniform occupation densities with a 2D Ornstein-Uhlenbeck process. Finally, we complete the paper with a conclusion and some perspectives. More details and all proofs are in the Appendices.

2. Multilevel Splitting Algorithm

2.1. Definition of the thresholds and related tools

In order to estimate the probability p that a particle starting from a point in some state space E reaches the critical subset $B \subset E$, we use the so-called "splitting algorithm" based on the nested sequence $B_1, \ldots, B_{M+1}$, defined in (1). Moreover, each frontier ∂B_k of B_k is partitioned into s disjoint subsets, denoted $\partial B_k^{(i)}$, such that:

$$\partial B_k = \bigcup_{i=1}^{s} \partial B_k^{(i)}, \quad k = 1, \ldots, M.$$

We assume that each ∂B_k has the same number s of subsets; this assumption is not restrictive as can be seen in the sequel. In any case, it is possible to rewrite the problem under concern in this particular setting.

The random dynamics of the particle are modelled by a stochastic process $Y = (Y_t; t \geqslant 0)$, and for $k = 1, \ldots, M+1$, we define τ_k as the first time that the particle hits ∂B_k. Hence, p can be written as $p = \mathbb{P}(\tau_{M+1} < \infty)$. For the sake of simplicity, we naturally assume that Y evolves continuously and that all of the intermediate thresholds are hit if the last threshold is hit. In fact, the dynamic under concern is not the particle dynamic directly, rather that of the embedded Markov chain observed at each time the particle hits a frontier ∂B_k. This embedded Markov chain will be denoted $(X_k)_{0\leqslant k\leqslant M+1}$. Thus, $X_k = i$ if the particle at time τ_k lies in $\partial B_k^{(i)}$, that is, $Y_{\tau_k} \in \partial B_k^{(i)}$.

Measures γ_k and functions f_k. We define for any $k = 1, \ldots, M$, a measure γ_k on the frontier ∂B_k by:

$$\gamma_k(i) = \mathbb{P}(X_k = i \ ; \ \tau_k < \infty).$$

This measure acts on the functions f defined on ∂B_k by $\gamma_k(f) = \mathbb{E}\left[f(X_k) \ ; \ \tau_k < \infty\right]$ in such a way that $\gamma_k(\mathbf{1}) = \mathbb{P}(\tau_k < \infty)$ is the probability that the particle hits the event B_k ($\mathbf{1}$ stands for the unit function).

For any $k = 1, \ldots, M$, we denote $\mathcal{M}_k$ (resp. $\mathcal{F}_k$) as the set of measures (resp. functions) defined on ∂B_k. In particular, the functions $f_k \in \mathcal{F}_k$ defined by:

$$f_k(i) = \mathbb{P}(\tau_{M+1} < \infty \mid X_k = i \ ; \ \tau_k < \infty), \quad k = 1, \ldots, M,$$

play a special role, since:

$$\gamma_k(f_k) = \sum_{i=1}^{s} \gamma_k(i) f_k(i) = p, \quad k = 1, \ldots, M. \tag{2}$$

In fact, $f_k(i)$ quantifies the difficulty of reaching the target set B starting from $\partial B_k^{(i)}$, while $\gamma_k(i) f_k(i)$ quantifies the difficulty of reaching B passing by $\partial B_k^{(i)}$, starting from O. Furthermore, for $k = 2, \ldots, M$, we introduce the operators P_k, $k = 2, \cdots, M$ defined on $\partial B_{k-1} \times \partial B_k$ by $P_k(i,j) = \mathbb{P}(X_k = j \ ; \ \tau_k < \infty | X_{k-1} = i \ ; \ \tau_{k-1} < \infty)$. Nevertheless, it is easier to consider P_k as an operator acting to the right on $\mathcal{F}_k$ as an operation $\mathcal{F}_k \to \mathcal{F}_{k-1}$ according to:

$$P_k(f)(i) = \mathbb{E}\left[f(X_k) \ ; \ \tau_k < \infty \mid X_{k-1} = i \ ; \ \tau_{k-1} < \infty\right],$$

and acting to the left on $\mathcal{M}_{k-1}$ as an operation $\mathcal{M}_{k-1} \to \mathcal{M}_k$ according to $(\mu P_k)(f) = \mu(P_k f)$. Each operator P_k is not Markovian, since the probability to reach ∂B_k is not equal to one; hence we define $g_{k-1} \in \mathcal{F}_{k-1}$ by:

$$g_{k-1}(i) \coloneqq P_k(\mathbf{1})(i) = \mathbb{P}(\tau_k < \infty \mid X_{k-1} = i \ ; \ \tau_{k-1} < \infty), \tag{3}$$

for $k = 2, \ldots, M$. Note that there is no need to define g_M since it would correspond to f_M.

We easily get the following transport relations for $k = 2, \ldots, M$:

$$\gamma_k = \gamma_{k-1} P_k, \quad f_{k-1} = P_k(f_k). \tag{4}$$

The notation is summarized in Figure 1.

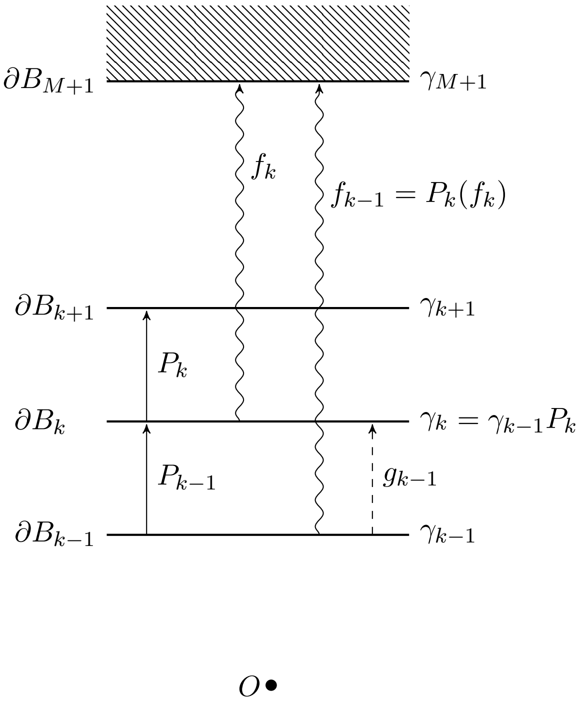

Figure 1: This figure summarizes the notation previously introduced.

Normalized measures μ_k. Since γ_k is not a probability measure, we define its normalized version μ_k on ∂B_k that acts on the functions $f \in \mathcal{F}_k$ in the following way:

$$\mu_k(f) := \frac{\gamma_k(f)}{\gamma_k(\mathbf{1})} = \mathbb{E}\left[f(X_k) \mid \tau_k < \infty\right],$$

(assuming that the thresholds have been chosen such that $\gamma_k(\mathbf{1}) \neq 0$ for all k). We notice that:

$$\mu_k(g_k) = \frac{\gamma_{k+1}(\mathbf{1})}{\gamma_k(\mathbf{1})} = \mathbb{P}(\tau_{k+1} < \infty \mid \tau_k < \infty), \tag{5}$$

and:

$$\mu_k(f_k) = \frac{p}{\gamma_k(\mathbf{1})} = \mathbb{P}(\tau_{M+1} < \infty \mid \tau_k < \infty).$$

Equation (4) induces the following scheme for the dynamics of μ_k:

$$\mu_k = \frac{\gamma_{k-1}(\mathbf{1})}{\gamma_k(\mathbf{1})} \mu_{k-1} P_k = \frac{1}{\mu_{k-1}(g_{k-1})} \mu_{k-1} P_k, \tag{6}$$

that leads to:

$$\gamma_k(f) = \mu_k(f) \gamma_1(\mathbf{1}) \prod_{p=1}^{k-1} \mu_p(g_p),$$

which applied successively to the functions f_k and f_{k+1} yields to $\mu_k(f_k) = \mu_{k+1}(f_{k+1}) \mu_k(g_k)$.

Convention. We extend the previous definitions to $k = 0$ and $k = M+1$. Considering that the particles are generated at the same point O, we define $\mathcal{F}_0$ as the set of constant functions and, in particular, $f_0 = p$ and $g_0 = \gamma_1(\mathbf{1})$. Analogously, $\mathcal{M}_0$ will represent the set of the Dirac measures at O up to a constant. Hence, $\gamma_0(f) = f$ (and $\mu_0 = \gamma_0$). Obviously, $\gamma_0(\mathbf{1}) = \mathbb{P}(\tau_0 < \infty) = 1$.

In the same way, B_{M+1} is reduced to a unique point, denoted by ω. Then, $\mathcal{F}_{M+1}$ is reduced to the constant functions, with $f_{M+1} = 1$ and $\mathcal{M}_{M+1}$ as the set of the Dirac measures at ω up to a constant, with $\gamma_{M+1}(f) = fp$ (such as $\gamma_{M+1}(f_{M+1}) = \gamma_{M+1}(\mathbf{1}) = p$) and $\mu_{M+1}(f) = f$. We also set $P_1(f) = \gamma_1(f)$ and $P_{M+1}(f) = f \times f_M$.

2.2. Multilevel Splitting Algorithm

To estimate the rare event probability, we proceed according to the algorithm already introduced in Glasserman et al. (1998) and Garvels (2000). Its principle is the following:

Initialization: We perform independently N particles from the same starting point O. A random number Z_1 of particles reach the threshold B_1, where Z_1 has a binomial distribution with parameters N and $\gamma_1(\mathbf{1})$. These Z_1 particles are spread over the subsets $\partial B_1^{(i)}$ according to a multinomial random variable (r.v.) $\text{Mult}(Z_1, \mu_1)$. Let $\mathbf{Z}_1$ be the corresponding random vector $(Z_{11}, \ldots, Z_{1r})$.

Step n $(2 \leqslant n \leqslant M)$**:** Each of the Z_{n-1} particles in ∂B_{n-1} is duplicated R_{n-1} times, so that a total number $R_{n-1}Z_{n-1}$ of particles is achieved. These new particles evolve according to the dynamics of the original process and the number Z_{nj} of particles reaching $\partial B_n^{(j)}$ is still a random number. Consider now the random vector $\mathbf{Z}_n = (Z_{n1}, \ldots, Z_{nr})$. The Z_{nj} particles in $B_n^{(j)}$ come from different subsets $\partial B_{n-1}^{(i)}$. Next, we decompose Z_{nj} in the following sum:

$$Z_{nj} = \sum_{i=1}^{s} Y_{nj}^{i}, \tag{7}$$

where Y_{nj}^i is the number of particles from $\partial B_{n-1}^{(i)}$ and having reached $\partial B_n^{(j)}$ whose total number $Y_n^i = \sum_{j=1}^s Y_{nj}^i$ is a binomial r.v. with parameters $R_{n-1}Z_{(n-1)i}$ and $g_{n-1}(i)$.

We represent the numbers Y_{nj}^i in a $s \times s$ tabular where each line $\mathbf{Y}_n^i = (Y_{n1}^i, \ldots, Y_{ns}^i)$, conditional on the knowledge of the total number Y_n^i, is distributed as a multinomial r.v. with parameters Y_n^i and $Q_n(i, \cdot)$ where:

$$Q_n(i, \cdot) := \frac{P_n(i, \cdot)}{g_{n-1}(i)} = \mathbb{P}(X_n = \cdot \mid X_{n-1} = i \; ; \; \tau_n < \infty).$$

$\cdot$	$\cdot$	$\cdot$	$\cdot$
$\cdot$	Y_{nj}^i	$\cdot$	Y_n^i
$\cdot$	$\cdot$	$\cdot$	$\cdot$
$\cdot$	Z_{nj}	$\cdot$	Z_n

Precisely, the random vector $\mathbf{Z}_n$ can be expressed as the sum $\mathbf{Z}_n = \sum_{i=1}^s \mathbf{Y}_n^i$ of the random vectors $\mathbf{Y}_n^i$, and the total number of particles at the end of step n is $Z_n = \sum_{i,j} Y_{nj}^i = \sum_{i=1}^s Y_n^i$.

Final step: Each of the Z_M particles in ∂B_M is duplicated R_M times to get a total number $R_M Z_M$ of particles. These new particles evolve according to the dynamics of the original process, and the Z_{M+1} particles, having reached ∂B_{M+1}, come from different subsets $\partial B_M^{(i)}$. We then decompose Z_{M+1} in the following sum:

$$Z_{M+1} = \sum_{i=1}^{s} Y_{M+1}^{i}, \tag{8}$$

where Y_{M+1}^i represents the number of particles from $\partial B_M^{(i)}$ and having reached ∂B_{M+1}. Conditional on the random vector $\mathbf{Z}_M$, the r.v.s Y_{M+1}^i, $i = 1, \ldots, r$ are independent and distributed as a binomial r.v. with parameters $R_M Z_{Mi}$ and $f_M(i)$. The result of this final step is simply the total number Z_{M+1} of particles in B_{M+1}, because B_{M+1} is being reduced to a point.

3. Algorithm analysis

In this section, we present a natural unbiased estimator of p and provide several expressions of its variance, including that found in Glasserman et al. (1996). We also define the cost of the algorithm.

3.1. A natural unbiased estimator of p

An estimator of the probability of hitting ∂B_{n+1}, conditional on ∂B_n having been hit, is naturally given by the ratio between the number of particles in ∂B_{n+1} and R_n times the number of particles in ∂B_n. From this, we deduce a natural estimator of the probability of interest p:

$$\widehat{p}_{M+1} = \frac{Z_1}{N} \times \prod_{n=1}^{M-1} \frac{Z_{n+1}}{R_n Z_n} \times \frac{Z_{M+1}}{R_M Z_M} = \frac{Z_{M+1}}{N R_1 \ldots R_M}. \tag{9}$$

Introducing the deterministic quantities $r_0 = N$ and $r_n = R_n r_{n-1}$, $n = 1, \ldots, M$, leads to $\widehat{p}_{M+1} = Z_{M+1}/r_M$. Then it is obvious to show that this estimator is unbiased. Indeed, by conditioning, (8) yields:

$$\mathbb{E}[\widehat{p}_{M+1}] = \frac{1}{r_{M-1}} \sum_{i=1}^{s} \mathbb{E}[Z_{Mi}] f_M(i).$$

To derive the mean of Z_{Mi}, notice that $\mathbb{E}[Z_{1j}] = N\gamma_1(j)$, $j = 1, \ldots, r$. By a new conditioning, we get that:

$$\mathbb{E}[Z_{2j}] = R_1 \sum_{i=1}^{s} \mathbb{E}[Z_{1i}] P_2(i,j) = N R_1 \gamma_2(j) = r_1 \gamma_2(j),$$

the last equality coming from (4). An induction principle allows us to establish that for $n = 2, \ldots, M$:

$$\mathbb{E}[Z_{nj}] = r_{n-1}\gamma_n(j), \tag{10}$$

that leads to $\mathbb{E}[\widehat{p}_{M+1}] = \sum_{j=1}^{s} \gamma_M(j) f_M(j) = \gamma_M(f_M) = p$.

3.2. The variance of the estimator

Proposition 3.1. *The coefficient of variation is given by:*

$$\frac{\mathrm{Var}(\widehat{p}_{M+1})}{p^2} = \sum_{k=1}^{M} \frac{1}{\gamma_k(\mathbf{1})}\left(\frac{1}{r_{k-1}} - \frac{1}{r_k}\right)\frac{\mathrm{Var}_{\mu_k}(f_k)}{\mathbb{E}^2_{\mu_k}(f_k)} + \sum_{k=0}^{M}\frac{1}{r_k\gamma_k(\mathbf{1})}\frac{1-\mu_k(g_k)}{\mu_k(g_k)}. \tag{11}$$

Introducing the operators Γ_{i+1} defined, for $f, g \in \mathcal{F}_{i+1}$, by $\Gamma_{i+1}(f,g) = P_{i+1}(fg) - P_{i+1}(f)P_{i+1}(g)$, we have:

$$\mathrm{Var}(\widehat{p}_{M+1}) = \sum_{i=0}^{M}\frac{1}{r_i}\gamma_i\left(\Gamma_{i+1}(f_{i+1})\right). \tag{12}$$

The variance is then split into two parts. The first sum outlines the variability due to the shape of the thresholds ∂B_k (defined by the f_k's), whereas the second outlines the variability due to the thresholds' number M, replication numbers R_k and thresholds position (contained in the P_k's and g_k's). Refer to Appendix A for more details on the operators Γ_{i+1}.

Comparison with other algorithms. Notice that for $s = 1$, the measures γ_k and the functions f_k are constant. Since $\mu_k(g_k) = \gamma_{k+1}/\gamma_k$, the expression of the variance becomes:

$$\frac{\mathrm{Var}(\widehat{p}_{M+1})}{p^2} = \sum_{k=0}^{M}\frac{1-\mu_k(g_k)}{r_k\gamma_{k+1}} = \sum_{k=0}^{M}\frac{1}{r_k}\left(\frac{1}{\gamma_{k+1}} - \frac{1}{\gamma_k}\right),$$

that corresponds to the expression established in Lagnoux-Renaudie (2008).

Furthermore, Formula (11) corresponds to Equation (2.21) established in Garvels (2000) for an algorithm with a single intermediate threshold. It has also been established in L'Ecuyer et al. (2009) and Cérou et al. (2011) in the general and continuous settings.

Finally, simple computation leads to the following expression:

$$\frac{\mathrm{Var}(\widehat{p}_{M+1})}{p^2} = \sum_{k=1}^{M}\frac{1}{\gamma_k(\mathbf{1})}\left(\frac{1}{r_{k-1}} - \frac{1}{r_k}\right)\frac{\mu_k(f_k^2)}{\mu_k^2(f_k)} + \left(\frac{1}{pr_M} - \frac{1}{r_0}\right),$$

that can be found in Glasserman et al. (1999).

3.3. The cost of the algorithm

The efficiency of the algorithm can be traduced in terms of the variance of the estimator that must be the smallest possible under the condition that the cost (in terms of computer time for example) remains finite. Our goal is then to derive the optimal parameters of the algorithm for a fixed cost.

The total number of particles generated during the algorithm is the r.v. $N + R_1Z_1 + \ldots + R_MZ_M$. From (10), $\mathbb{E}[Z_n] = r_{n-1}\gamma_n(\mathbf{1})$, the mean of the total number of particles generated by the algorithm is:

$$C^{(0)}_{M+1} \coloneqq r_0 + r_1\gamma_1(\mathbf{1}) + \ldots + r_M\gamma_M(\mathbf{1})$$

and can be considered as a natural cost.

Now we present a more realistic cost that takes into account the probability of $P_k(i,j)$ to reach $\partial B_k^{(j)}$ from $\partial B_k^{(i)}$. In fact, even if the algorithm presented here is based on the simulation of multinomial r.v.s, the introduction of this new cost allows us to consider the dynamics of a particle between two successive thresholds through the functions g_k. Thus, we associate with each particle from $\partial B_k^{(i)}$ a unitary cost $c_k(i)$ that depends on the starting threshold and the difficulty $g_k(i)$ to succeed in reaching the next threshold. More precisely, we assume that:

$$c_0 = c(\gamma_1(\mathbf{1})), \quad c_k(i) = c(g_k(i)), \quad k = 1, \ldots, M,$$

where c is a positive function, that is decreasing (the smaller the probability of success is the highest the cost is), such that $c(x)$ converges to a constant (in general small) when x tends to 1.

Proposition 3.2. *The mean cost is given by:*

$$C_{M+1} = Nc_0 + \sum_{n=1}^{M} r_n \sum_{i=1}^{s} \gamma_n(i)c_n(i) = \sum_{n=0}^{M} r_n \gamma_n(c_n). \tag{13}$$

The approach presented here leads to a relatively simple formula for the total mean cost, similar to that used in Lagnoux-Renaudie (2008). The multidimensionality of the model is taken into account through the function c_n.

4. Algorithm optimisation

Before proceeding to the optimisation of the algorithm, we revisit the general setting.

4.1. General setting

In many applications, the rare event probability p can be viewed as an overflow probability. More precisely, let h be a real-valued measurable function defined on E, and let $L \geqslant 0$ be a given threshold. Then, p is rewritten as $p = \mathbb{P}(h(Y_t) \geqslant L)$, where the process Y has been defined in the Introduction. As a consequence, we can naturally use the function h to determine the intermediate thresholds and apply the splitting methodology to the real-valued process Z defined by $Z_t := h(Y_t)$, for all $t \geqslant 0$ (for simplicity $Z_0 \geqslant 0$). For the sake of simplicity, we assume that Z evolves continuously and that all of the intermediate thresholds are hit if the last threshold is hit.

However, note that the intermediate thresholds $L_1, \ldots, L_{M+1}$ for Z define splitting surfaces $\partial B_1, \ldots, \partial B_{M+1}$ for Y by $\partial B_k = \{y \in E \mid h(y) = L_k\}$. Defining the levels ∂B_k in this way is not well adapted and is far from optimal. Indeed, this methodology is geometrical, and is only based on a level set without taking into account the probabilistic aspects. More precisely, it seems natural to incorporate information regarding the difficulty of reaching the target set from any point of the ∂B_k. This information is precisely given by the function f_k, introduced previously. Assuming the possibility to define a function f, (the importance function), on the whole space E by:

$$f(x) = \mathbb{P}(\tau_{M+1} < \infty \mid \text{starting from } x),$$

rather, we define ∂B_k as the set of points $x \in E$, such that $f(x) = L_k$ for some $L_k \in [0, 1]$. In some sense, we use iso-probability density levels as intermediate thresholds. Of course, the difficulty here is to determine the function f, the thresholds number M and the values of L_k. Nevertheless, there exist methods that allow us to get estimators of f using a reverse time analysis, as proposed in Garvels (2000).

To illustrate the importance of a good choice of the intermediate thresholds, let us consider the following example represented in Figure 2.

Example 4.1. *With $M = 1$ and a threshold ∂B_1 partitioned in two subsets such that:*

$$\begin{cases} \gamma_1(1) = 10^{-2}, \ \gamma_1(2) = 0.5, \\ f_1(\mathbf{1}) = 10^{-1}, \ f_1(2) = 10^{-3}, \end{cases}$$

we obtain $p = 1.5 \cdot 10^{-3}$ and $\gamma_1(\boldsymbol{1}) = 0.51$.
Let us simulate particles starting from O. We expect that 51% of particles will reach the threshold ∂B_1, with 50% in $\partial B_1^{(2)}$ and only 1% in $\partial B_1^{(1)}$. Nevertheless among those in ∂B_1, the particles in the subset $\partial B_1^{(1)}$ are 100 times more likely to reach the target set than those in $\partial B_1^{(2)}$. Thus, using this design of ∂B_1 leads to the pointless simulation of almost 50% of particles. We see that using a function $f_1 = p/\alpha$, with $\alpha \in]0, 1[$, implies that $\gamma(\boldsymbol{1}) = \alpha$, and all the particles in ∂B_1 then have the same probability to reach the target set.

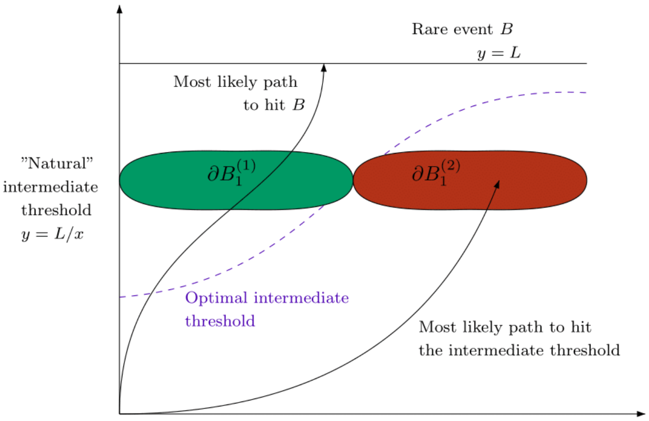

Figure 2: The crucial choice of the importance function using an example.

The construction of the importance function when the target probability has a large deviation characterization is discussed in Dean and Dupuis (2009). This context is also considered in Sadowsky (1996) and Remark 4.2. Nevertheless, it seems difficult to translate the results obtained into the framework of this paper.

4.2. Optimisation

It is important to keep in mind Equation (11) and the fact that the variance of the $\widehat{p}_{M+1}$ can be split in two parts: the first resuming the variability due to the shape of the thresholds, and the second resuming the variability due to the thresholds' number, replication numbers and thresholds' position.

Furthermore, the splitting algorithm's parameters are the initial number N of particles, the replication numbers $R_1, \ldots, R_M$, the number M of intermediate thresholds, and their characteristics (through the P_k's and the g_k's).

Proposition 4.1. *The parameters of the algorithm optimised by minimisation of the variance of the estimator for a fixed cost are the following:*

(i) the functions f_k so that they do not depend on the starting point in ∂B_k; and

(ii) *the optimal values of the parameters N, M, $\{R_k\}_{k=1}^{M}$ and $\{P_k\}_{k=1}^{M+1}$ obtained in Lagnoux-Renaudie (2008) for the unidimensional case (i.e. $s = 1$). More precisely, N is related to C_{M+1} and all the R_k's are equal to the same value, say R, which depends on N and C_{M+1}. Furthermore, in order to satisfy the tradeoff between a premature stop of the algorithm ($R_kP_{k+1} \ll 1$) and a prohibitive cost ($R_kP_{k+1} \gg 1$), we need the condition $R_kP_{k+1} = 1$. Then, M is fixed by the relation $Rp^{1/(M+1)} = 1$.*

As expected, the optimal choice consists of taking the thresholds ∂B_k in such a way that f_k is constant. This is consistent with the observations of Section 4.1. Nevertheless, the difficulty lies in the evaluation of the importance function f, and consequently, in the design of the thresholds. We will see in Section 6 the impact of a non-optimal choice on the variance and on the cost of the algorithm.

If for some k, the function f_k is constant, given that $\gamma_k(f_k) = p$, we get the following identity $f_k = p/\gamma_k(\mathbf{1})$. Moreover, it comes from the definition of μ_k and Equations (A.2) and (5) that $\Gamma_k(f_k) = g_{k-1}(1-g_{k-1})p^2/\gamma_k^2(\mathbf{1})$ and:

$$\gamma_{k-1}\left(\Gamma_k(f_k)\right) = \frac{p^2}{\gamma_k(\mathbf{1})}\frac{\mu_{k-1}\left(g_{k-1}(1-g_{k-1})\right)}{\mu_{k-1}(g_{k-1})}.$$

Additionally, by (4), the function f_{k-1} can be expressed as $f_{k-1} = g_{k-1}p/\gamma_k(\mathbf{1})$. Moreover, if f_{k-2} is also constant, then $P_{k-1}(g_{k-1}) = \gamma_k(\mathbf{1})/\gamma_{k-2}(\mathbf{1})$, and after calculus:

$$\gamma_{k-2}\left(\Gamma_{k-1}(f_{k-1})\right) = p^2\left[\frac{1}{\gamma_{k-1}(\mathbf{1})}\frac{\mu_{k-1}(G_{k-1}^2)}{\mu_{k-1}^2(g_{k-1})} - \frac{1}{\gamma_{k-2}(\mathbf{1})}\right].$$

Finally, if all the functions f_k are constant, then the functions g_k are also constant: $g_k = \gamma_{k+1}(\mathbf{1})/\gamma_k(\mathbf{1})$, which is also the case for the functions $\Gamma_{k+1}(f_{k+1})$:

$$\Gamma_{k+1}(f_{k+1}) = \frac{p^2}{\gamma_k(\mathbf{1})}\left[\frac{1}{\gamma_{k+1}(\mathbf{1})} - \frac{1}{\gamma_k(\mathbf{1})}\right].$$

Remark 4.2. *These results justify the choices made in the algorithm proposed in Miretskiy et al. (2009). The authors assume that:*

$$\lim_{B\to\infty}\frac{1}{B}\log p_B^s = -\gamma(s) \quad \forall s \notin A;$$

where p_B^s represents the probability of reaching the target event A starting from s, with B the rarity parameter, and γ a decreasing function. The algorithm consists of taking:

- *the replication numbers (except for the last one) all equal to R;*
- *the number of thresholds n_B equal to $\lfloor B\gamma(s)/\log(R)\rfloor$;*
- *the frontier l_k of the intermediate threshold L_k equal to*

$$\left\{x \in D \quad / \quad \gamma(s) - \gamma(x) = \frac{k}{B}\log R\right\} \qquad k = 0\ldots n_B;\ \text{and}$$

- *the last replication number equal to $R' = \left\lfloor e^{B\gamma(s)-n_B\log R}\right\rfloor$.*

In other words, the authors equal all the replication numbers (except, eventually, the last one), then, take the number of thresholds equal to the optimal one in Lagnoux-Renaudie (2008). Finally, they fix all the thresholds in such a way that the decreasing rate $\gamma(s)$ is uniform over the thresholds, and the probability of reaching the target set A starting from the k-th threshold depends on k but not on the starting point of the frontier l_k.

5. Sensitivity analysis: deletion of a threshold

Now, we study the sensitivity of $\mathrm{Var}(\widehat{p}_{M+1})$ with respect to the number of thresholds. We assume that the thresholds have the optimal shape: the functions f_k are constant. This works in the unidimensional setting. Optimally, the thresholds are such that all of the transition probabilities are equal, but as p is unknown, this value cannot be computed. Moreover, in practice, the freedom of the choice of the thresholds can be limited by physical constraints. Next, we study the need for an intermediate threshold and derive a procedure to detect whether or not we should keep it.

5.1. Iterative expressions of variance and cost

The goal of this section is to compare the variance and the cost of the estimator obtained with M thresholds with those obtained in the same setting, but deleting the k-th threshold (thus, in a simulation with $(M-1)$ thresholds). We reallocate the replication numbers as follows:

- for any $j = 1, \dots, k-2$, R_j stays unchanged;
- R_{k-1} is replaced by $\lambda_{k-1} R_{k-1} R_k$; and
- for any $j = k, \dots, M-1$, R_j is modified in $\lambda_j R_{j+1}$.

For instance, we can decide to keep all of the R_j's unchanged so the replication numbers are $R_1, \dots, R_{k-1},\ R_{k+1} \dots, R_M$, or to report the replication number of the k-th threshold on the $k-1$-th's, the replication numbers being $R_1, \dots, R_{k-1}R_k, R_{k+1}, \dots, R_M$.

Proposition 5.1. *The variance of the estimator $\widehat{p}_{M+1}$ with M thresholds is the sum of the variance of the estimator $\widehat{p}_M^{(-k)}$ obtained by running the algorithm with the k-th threshold deleted (thus with $(M-1)$ intermediate thresholds) and the contribution of the k-th threshold:*

$$\begin{aligned}\mathrm{Var}\left(\widehat{p}_{M+1}\right) = \mathrm{Var}\left(\widehat{p}_M^{(-k)}\right) &+ \frac{1}{r_{k-1}}\left(1 - \frac{1}{\Lambda_{k-1}R_k}\right)\gamma_{k-1}\left(\Gamma_k(f_k)\right) \\ &+ \sum_{j=k}^{M} \frac{1}{r_j}\left(1 - \frac{1}{\Lambda_{j-1}}\right)\gamma_j\left(\Gamma_{j+1}(f_{j+1})\right),\end{aligned}$$

where $\Lambda_p = \prod_{j=k-1}^{p} \lambda_j$.

Similarly, the cost C_{M+1} given in (13) is the sum of the cost $C_M^{(-k)}$, computed with $M-1$ intermediate thresholds, and the contribution of the k-th threshold:

$$C_{M+1} = C_M^{(-k)} + r_{k-1}\left[\gamma_{k-1}(c_{k-1}) - R_k\Lambda_{k-1}\gamma_{k-1}(\tilde{c}_{k-1})\right] + r_k\gamma_k(c_k) + \sum_{j=k+1}^{M} r_j\gamma_j(c_j)\left(1 - \Lambda_{j-1}\right),$$

where $\tilde{c}_{k-1}$ stands for the cost of a particle going from the $(k-1)$-th threshold to the k-th in an algorithm with $(M-1)$ levels.

The free parameters of the new algorithm with $M-1$ intermediate thresholds are $\{\Lambda_{j-1}\}_{j=k}^{M}$: these can be chosen by keeping the cost constant. It is sufficient to take:

$$\Lambda_{k-1} = \frac{\gamma_{k-1}(c_{k-1}) + R_k\gamma_k(c_k)}{R_k\gamma_{k-1}(\tilde{c}_{k-1})} \quad \text{and} \quad \Lambda_{j-1} = 1,\ j = k+1, \dots, M. \tag{14}$$

With these values, the variance $\mathrm{Var}\left(\widehat{p}_{M+1}\right)$ becomes:

$$\mathrm{Var}\left(\widehat{p}_M^{(-k)}\right) + \frac{1}{r_{k-1}}\left(1 - \frac{1}{\Lambda_{k-1}R_k}\right)\gamma_{k-1}\left(\Gamma_k(f_k)\right) + \frac{1}{r_k}\left(1 - \frac{1}{\Lambda_{k-1}}\right)\gamma_k\left(\Gamma_{k+1}(f_{k+1})\right).$$

5.2. Is the k-th threshold useful?

The goal is to study the need for an intermediate threshold and derive a procedure to detect whether or not we should keep it. More precisely, the k-th threshold will be deleted if the variance of $\widehat{p}_M^{(-k)}$ is lower than the one of $\widehat{p}_{M+1}$, that is, if the contribution of the k-th threshold is positive. In order to get a tractable procedure, we assume the following:

(A1) All thresholds have the optimal shape. Then, we find a unidimensional algorithm ($s = 1$), so the measures γ_j and the functions f_j, g_j are constant; and

(A2) The cost $\tilde{c}_{k-1}$ between ∂B_{k-1} and ∂B_{k+1} in the algorithm without the k-th threshold is given by $\tilde{c}_{k-1} = c_{k-1} + c_k$. Notice that each $c_k = c(g_k)$ is constant by (A1).

With these assumptions, we get:

$$\Lambda_{k-1} = \frac{a_k}{R_k} + g_{k-1}(1 - a_k) \quad \text{and} \quad \Lambda_j = 1,\ j = k, \dots, M-1,$$

where $a_k \coloneqq c_{k-1}/(c_{k-1} + c_k)$. Now, plugging these values into the variance, we get:

$$\mathrm{Var}(\widehat{p}_{M+1}) = \mathrm{Var}(\widehat{p}_M^{(-k)}) + \frac{p^2 Q(g_{k-1})}{r_k \gamma_{k+1} g_{k-1}\left[a_k + R_k g_{k-1}(1 - a_k)\right]},$$

where:

$$Q(x) = -x^2 R_k(R_k\beta - 1)(1 - a_k) + x\left[R_k(R_k\beta - 1)(1 - a_k) - a_k(R_k - 1)\right] + (R_k - 1)\beta a_k,$$

and β is defined by:

$$\beta \coloneqq g_{k-1} g_k = \frac{\gamma_{k+1}}{\gamma_{k-1}} \in [0, g_{k-1}]. \tag{15}$$

Notice that $\beta = \mathbb{P}(\tau_{k+1} < \infty \mid \tau_{k-1} < \infty)$ quantifies the level of difficulty for a particle to go from ∂B_{k-1} to ∂B_{k+1} and so β does not depend on the deleted k-th threshold.

The sign of Q in the corrective term is the opposite of that in the following polynomial:

$$R(x) = x^2 - (1 - \alpha)x - \alpha\beta, \qquad \text{with} \qquad \alpha = \frac{a_k(R_k - 1)}{R_k(1 - a_k)(R_k\beta - 1)},$$

at $x = g_{k-1} \in]0, 1[$. Its discriminant is $\Delta = (1 - \alpha)^2 + 4\alpha\beta$.

In practice, we start by realising a pre-run in order to estimate the unknown parameters γ_{k-1} and γ_k, and thus g_{k-1} and β. Then, the procedure is the following:

1. If $R_k\beta = 1$: $Q(g_{k-1}) = a_k(R_k - 1)(\beta - g_{k-1}) \leqslant 0$ and it is recommended to preserve the k-th threshold;
2. If $R_k\beta > 1$: Δ is strictly positive and R has two roots of opposite signs, $x_k^- < 0 < x_k^+ < 1$:
   (a) when $0 < g_{k-1} < x_k^+$, the polynomial Q is positive and it is recommended to delete the k-th threshold; and
   (b) when $x_k^+ < g_{k-1} < 1$, the polynomial Q is negative and it is recommended to preserve the k-th threshold;
3. If $R_k\beta < 1$ and $\Delta < 0$: the polynomials R and Q are positive and it is recommended to delete the k-th threshold; and
4. If $R_k\beta < 1$ and $\Delta > 0$: the polynomial R has two roots $x_k^- < x_k^+$:

(a) when $0 < g_{k-1} < x_k^-$, the polynomials R and Q are positive and it is recommended to delete the k-th threshold; and
(b) if $x_k^- < 1$, when $x_k^- < g_{k-1} < 1$, the polynomials R and Q are negative and it is recommended to preserve the k-th threshold.

Next, we focus on the simplified cost as analytical values may be obtained.

Proposition 5.2. *Considering the simplified cost, there is no reason to introduce a new threshold when $\beta \geqslant 1/9$. When $\beta < 1/9$, the optimal positioning minimising the variance for a fixed cost is given by $g_{k-1} = (1 - 3\beta)/2$. In that case, the optimal replication number is:*

$$R_k^* = \frac{2(1-5\beta)}{1-9\beta^2}\left(1 + \sqrt{\frac{2(1-\beta)}{1-5\beta}}\right),$$

which decreases from $2(1+\sqrt{2})$ for $\beta = 0$ to 3 for $\beta = 1/9$.

6. Sensitivity analysis: perturbation of a threshold

In this section, we assume that all of the thresholds ∂B_i are optimal (i.e. f_i constant), except ∂B_k. Thus:

$$f_k = \frac{p}{\gamma_{k+1}(\mathbf{1})} g_k, \quad P_k(g_k) = \mu_k(g_k)\mu_{k-1}(g_{k-1}),$$

and the variance is given by:

$$\frac{\mathrm{Var}(\widehat{p}_{M+1})}{p^2} = \frac{1}{\gamma_k(\mathbf{1})}\left(\frac{1}{r_{k-1}} - \frac{1}{r_k}\right)\frac{\mathrm{Var}_{\mu_k}(f_k)}{\mathbb{E}^2_{\mu_k}(f_k)} + \sum_{i=0}^{M} \frac{1}{r_i\gamma_i(\mathbf{1})}\frac{1-\mu_i(g_i)}{\mu_i(g_i)}.$$

With a pre-run of the algorithm, we estimate the values of $g_k(i)$ for $i = 1, \ldots, r$, and thus, $\mu_k(g_k)$.

We now want to twist ∂B_k in order to get closer to the optimal shape and to obtain a new function f_k constant. Consequently, with this new threshold, all of the functions f_k become constant, as does the new function g_k, as explained in Section 4.2. Introducing the new threshold $\partial \widetilde{B}_k$ implies that γ_k, P_k, P_{k+1}, g_k and g_{k-1} are changed accordingly, and we use a $\widetilde{}$ symbol to denote the new terms.

Furthermore, in order to guarantee a slight perturbation of threshold k, we assume naturally that:

$$B_{k+1} \subset \widetilde{B}_k \subset B_{k-1},$$

which implies that $\gamma_{k+1}(1) \leqslant \widetilde{\gamma}_k(1) \leqslant \gamma_{k-1}(1)$. We also introduce two operators E_k and E_{k+1} defined by:

$$\widetilde{P}_k = P_k E_k, \quad \widetilde{P}_{k+1} = E_{k+1}P_{k+1},$$

such that $\widetilde{P}_k\widetilde{P}_{k+1} = P_kP_{k+1}$. So defined, E_k (respectively E_{k+1}) is an operator acting on $\widetilde{\mathcal{F}}_k$ (resp. $\mathcal{F}_k$) valued in $\mathcal{F}_k$ (resp. $\widetilde{\mathcal{F}}_k$). We have:

$$\widetilde{g}_{k-1} = \widetilde{P}_k(1) = P_k(E_k 1) \quad \text{and} \quad \widetilde{g}_k = \widetilde{P}_{k+1}(1) = E_{k+1}(g_k).$$

Note that $\widetilde{g}_{k-1}$ and $\widetilde{g}_k$ are constant and linked by the identity $\beta = \widetilde{g}_{k-1}\widetilde{g}_k$. If we choose $E_{k+1}(i,j) = \delta_{ij}/a_i$ with $a_i = Kg_k(i)$ for some constant K, then $E_{k+1}(g_k) = 1/K$ so that $\widetilde{g}_k = 1/K$. Moreover, since $\widetilde{g}_k = \gamma_{k+1}(\mathbf{1})/\widetilde{\gamma}_k(\mathbf{1})$, we get:

$$K = \frac{\widetilde{\gamma}_k(\mathbf{1})}{\gamma_{k+1}(\mathbf{1})}. \tag{16}$$

Furthermore, taking $E_k(i,j) = Kg_k(j)\delta_{ij}$ leads to $E_kE_{k+1} = Id$ (and we recover $\widetilde{P}_k\widetilde{P}_{k+1} = P_kP_{k+1}$). Finally:

$$\widetilde{P}_{k+1}(i,j) = \frac{1}{K}\frac{P_{k+1}(i,j)}{g_k(i)} \quad \text{and} \quad \widetilde{P}_k(i,j) = Kg_k(j)P_k(i,j).$$

As a consequence, if $Kg_k(i) > 1$, $\widetilde{P}_{k+1}(i,j) < P_{k+1}(i,j)$ for any j and $\widetilde{P}_k(l,i) < P_k(l,i)$ for any l. It remains to determine the optimal value of K by keeping the total cost of the algorithm constant, which translates as:

$$\widetilde{c}_{k-1}\gamma_{k-1}(\mathbf{1}) + R_k\widetilde{c}_kK\gamma_k(g_k) = \gamma_{k-1}(c_{k-1}) + R_k\gamma_k(c_k),$$

leading to:

$$K = \frac{1}{\widetilde{c}_k}\left\{\frac{\mu_k(c_k)}{\mu_k(g_k)} + \frac{1}{R_k\beta}[\mu_{k-1}(c_{k-1}) - \widetilde{c}_{k-1}]\right\}.$$

Notice that fixing the value of K amounts to defining the value of $\widetilde{\gamma}_k(\mathbf{1})$ by equation (16). Also note that if the cost function c is constant and equal to 1, then the optimal value of K reduces to $K = 1/\mu_k(g_k)$.

Numerical application. Considering a two-dimensional Ornstein-Uhlenbeck process, we illustrate a way by which to deform the shape of a threshold in order to obtain iso-probability levels. To this end, we simulate the stochastic process defined by:

$$\begin{cases} \mathrm{d}X_t = -\Lambda X_t\,\mathrm{d}t + \sigma\,\mathrm{d}W_t,\ t > 0 \\ X_0 = x \in \mathbb{R}^2 \end{cases}$$

where $\Lambda = \mathrm{diag}(\lambda_1, \lambda_2)$ with $\lambda_1 > \lambda_2 > 0$, $\sigma > 0$ and W is a two-dimensional standard Brownian motion.

We start the algorithm by independently generating $N = 300$ particles from $x = (0.05, 0)$, and consider the 0.5 radius circle as the first intermediate threshold ∂B_1. In the sequel, we take $M = 2$, $B_2 = D(0,1)$ and $B = B_{M+1} = B_3 = D(0,1.5)$. The parameters of the stochastic process are $\lambda_1 = 1, \lambda_2 = 0.2, \sigma = 0.3$ and its simulation is completed via an Euler Scheme with a step of 0.01 (we use the software Mathematica Wolfram Research (2015)). First, we estimate the density of the occupancy measure of the process on ∂B_1[1], with respect to its related Lebesgue measure. This estimation is based on the von Mises Kernel, and, as expected, this density (represented in Figure 3 left-hand side) is far from uniform.

We previously noted that the efficiency of the splitting algorithm is be enhanced when the occupancy measures of the process on the intermediate thresholds are uniform. In our case,

[1] Since we work with continuous processes, the particles evolve until they reach ∂B_1 or the small disk $D(0, 0.01)$ instead of the origin.

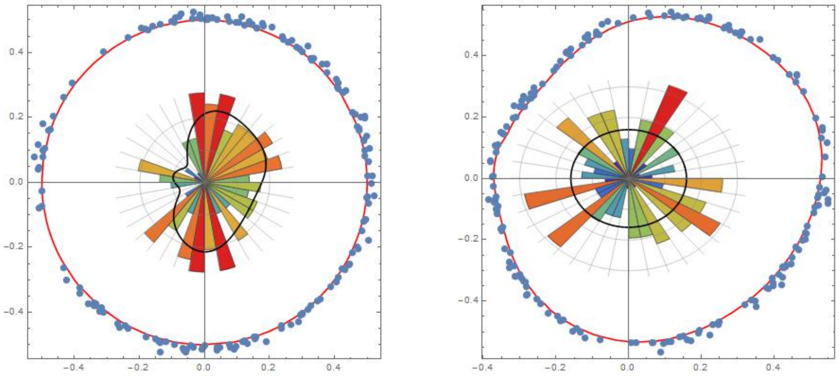

Figure 3: The density of the occupancy measure at the first intermediate threshold and its estimation based on the von Mises kernel (black line). On the left-hand side, the threshold is the centered 0.5 radius circle whereas on the right-hand side, the threshold is the conformal image of the circle.

since λ_1 is greater than λ_2, we presume that the suited thresholds are ellipses. This intuition is confirmed by Figure 3 (left-hand side) and is consistent with Theorem (1.3) in Antonini (1991) that establishes that, for any given $x \in \mathbb{R}^2$, $(Z_t)_{t\geqslant 0} := \left(\frac{\sqrt{2}}{\sigma\sqrt{\log t}} X_t\right)_{t\geqslant 0}$ admits the ellipse $\mathcal{E} = \{y = (y_1, y_2) \in \mathbb{R}^2;\ \lambda_1 y_1^2 + \lambda_2 y_2^2 \leqslant 2\}$ while t goes to infinity.

Second, our goal is to deform the first threshold. As the process lives in the plane, we can use a conformal map, $\varphi_1 : B_1 \to \Omega_1$, in order to obtain a uniform occupancy distribution. Notice that the conformal maps are very convenient as planar transformations since they allow only local rotations and scales avoiding disturbing distortions. Particularly, for common domains, $\partial\Omega_1 = \varphi_1(\partial B_1)$. We follow the procedure described in Weber and Gotsman (2010) to construct the conformal map (see also Appendix C for more details). Once the conformal map φ_1 is computed (see Figure 3 right-hand side), we restart the algorithm using the threshold $\partial\Omega_1$ instead of ∂B_1. Then, we start the next step by first estimating the density of the occupation measure on ∂B_2 after duplication of the particles in $\partial\Omega_1$, and second, deforming the shape of ∂B_2, as previously. Once $\partial\Omega_2$ is obtained, we restart the algorithm with this new threshold. Finally, the final step estimates the conditional probability to reach ∂B_3 for the particles in $\partial\Omega_2$. More precisely:

1. Each of the particles in $\partial\Omega_1$ is duplicated $R_1 = 2$ times and evolve independently from $\partial\Omega_1$ until $D(0, 0.01)$ or ∂B_2 is reached. We determine the density of the occupancy measure of the process on ∂B_2 by using the von Mises kernel. Then we find a conformal map $\varphi_2 : B_2 \to \Omega_2$ such that $\varphi_2(\partial B_2) = \partial\Omega_2$ and the image of the occupancy measure on ∂B_2 is the uniform measure on $\partial\Omega_2$ (see Figure 4 for more details);
2. We independently perform a second set of particles from their same starting point at $\partial\Omega_1$ with the same size, and stop them as soon as $D(0, 0.01)$ or $\partial\Omega_2$ is reached; and
3. Each of the particles in $\partial\Omega_2$ is duplicated $R_2 = 2$ times, and evolve independently from $\partial\Omega_2$ until $D(0, 0.01)$ or ∂B_3 is reached.

We emphasize that our intent is not to propose a new algorithm based on conformal mappings since we do not have sufficient expertise to produce an efficient code. Working in 2D is already difficult, and considering a greater dimension becomes even more complex. Nonetheless, the harmonic functions or the quasi-conformal maps (see Ahlfors (2006) and Heinonen (2006)) are the natural generalization of the conformal transformations in higher dimensions. In our particular context, we start by estimating the density of occupation probability on a sphere, which

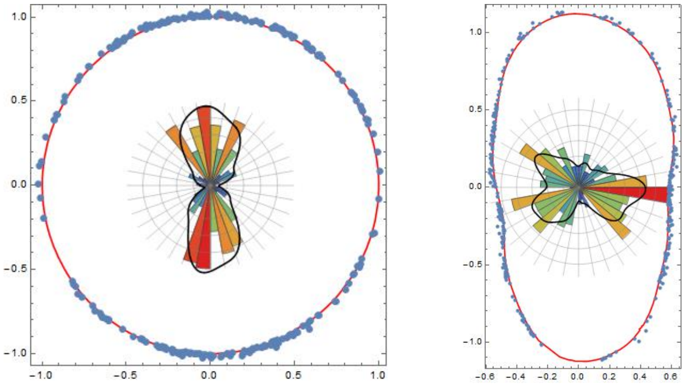

Figure 4: Using a replication factor $R_1 = 2$ for the particles having reached the first deformed threshold, we make these particles evolve until they reach the next threshold or the inner 0.01 radius circle. The empirical densities of the occupancy measure of the unit circle (left) and of the deformed threshold (right) and their respective estimations based on the von Mises kernel (black line) are represented. As mentioned before, we generate a new set of particles from the first deformed threshold instead of keeping the particles used to determine the conformal map. Thus in any rigor, we do not precisely recover the image measure; which explains the relative gap from the uniform distribution.

we can be identified as a volume form. Therefore, we can attempt to determine a Riemannian metric g such that associated Riemannian volume form is equal to the previous one. Then, we can deform the metric g into the uniform metric through a Ricci flow, for instance; we finally get a new Riemannian variety homeomorphic to the sphere. For more details, see for example, Patane et al. (2014).

7. Conclusion and Perspectives

In this paper, we continue the multidimensional approach studied in Glasserman et al. (1998) and Garvels (2000) in order to obtain a new expression of the variance of the estimator analogous to that of the continuous case (L'Ecuyer et al. (2009)). Then, we derive the optimal parameters of the splitting algorithm. Furthermore, by introducing new operators, we obtain alternative expressions of the variance which are more tractable when we compare the variance of the estimators in an algorithm with M thresholds with an algorithm in which one of the threshold has been deleted. More precisely, we derive a procedure in order to detect whether or not we keep it. Finally, we investigate the sensitivity of the variance of the estimator with respect to a deviation of the threshold shape from the optimal one. We illustrate our theoretical results using the planar Ornstein-Uhlenbeck process, for which we propose a procedure based on conformal maps to twist the thresholds in order to get closer to the optimal shapes.

A next natural research direction is the creation of a new algorithm that can decide the thresholds on the fly, for instance, by using an efficient algorithm for shape deformation. Such an algorithm has been proposed in Cérou and Guyader (2005) but applies only to 1D frameworks. When working in 2D or more, the problem is even more complex and challenging; see for instance, Dean and Dupuis (2009) for an approach based on the sub-solutions of the Hamilton-Jacobi-Bellman equation. Another way to investigate would be by using a test example to compare different existing algorithms dedicated to 2D or more, such as subset simulation (Au and Beck (2001)), importance sampling based on design points (Au et al. (1999),and Kiureghian and Dakessian (1998)), or adaptive pre-samples (Au and Beck (1999)).

Acknowledgement

The authors are greatly indebted to the referees for their fruitful and detailed suggestions or comments which have permitted us to significantly improve our paper.

Appendix A. The operator Γ_k and its iterates

The variance of the estimator of the target probability involves operators Γ_k and their iterates defined in the following way. For $k = 1, \dots, M+1$, we introduce Γ_k acting on $\mathcal{F}_k \times \mathcal{F}_k$ and valued in $\mathcal{F}_{k-1}$ by, for $f,\, g \in \mathcal{F}_k$,

$$\Gamma_k(f,g) = P_k(fg) - P_k(f)P_k(g). \tag{A.1}$$

To lighten notation, we denote $\Gamma_k(f)$ for $\Gamma_k(f,f)$. With the previous notation, one has:

$$\Gamma_1(f,g) = \gamma_1(fg) - \gamma_1(f)\gamma_1(g) \quad \text{ and } \quad \Gamma_{M+1}(f,g) = fg \times f_{M+1}(1 - f_{M+1}) = fg.$$

Firstly, we straightforwardly check that Γ_k is bilinear and symmetric. Secondly, writing $\Gamma_k(f)(i)$ as a conditional variance, we get:

$$\Gamma_k(f)(i) = P_k(f^2)(i) - P_k(f)(i)^2 = \mathbb{E}^c\left(\left[f(X_k)\mathbf{1}_{\{\tau_k<\infty\}} - \mathbb{E}^c\left(f(X_k)\mathbf{1}_{\{\tau_k<\infty\}}\right)\right]^2\right)$$

where $\mathbb{E}^c$ is the expectation conditionally to the set $\{X_{k-1} = i;\ \tau_{k-1} < \infty\}$, we get $\Gamma_k(f) \geqslant 0$. Moreover, by (3), $\Gamma_k(f,\mathbf{1}) = (1 - g_{k-1})P_k(f)$, $\quad \Gamma_k(\mathbf{1}) = g_{k-1}(1 - g_{k-1})$ and by (4):

$$\Gamma_k(f_k) = P_k(f_k^2) - f_{k-1}^2. \tag{A.2}$$

Now, let us iterate the construction of Γ_k. In that view, we introduce the multiplicative operator $\Gamma_k^{(0)}$ defined by:

$$\Gamma_k^{(0)}(f,g) = fg, \quad f, g \in \mathcal{F}_k,$$

such that $\Gamma_k(f,g) = P_k(\Gamma_k^{(0)}(f,g)) - \Gamma_{k-1}^{(0)}(P_k(f), P_k(g))$. This suggests that we define for any k, the iterated operators $\Gamma_k^{(n)}$ in the following way:

$$\Gamma_k^{(n+1)}(f,g) = P_{k-n}\left(\Gamma_k^{(n)}(f,g)\right) - \Gamma_{k-1}^{(n)}(P_k(f), P_k(g)), \quad k \geqslant 1,\ 0 \leqslant n \leqslant k-1 \tag{A.3}$$

with the convention $P_1(f) = \gamma_1(f)$ outlined above. The operator $\Gamma_k^{(n)}$ valued in $\mathcal{F}_{k-n}$ acts on $\mathcal{F}_k \times \mathcal{F}_k$.

We use the same simplified notation $\Gamma_k^{(n)}(f)$ to refer to $\Gamma_k^{(n)}(f,f)$. We introduce $P_{p,n}$ defined by:

$$\begin{aligned} P_{n,n} &= Id, \quad n = 0, \dots, M, \\ P_{p,n} &= P_{p+1} \dots P_n, \quad 0 \leqslant p \leqslant n-1, \quad n = 1, \dots, M. \end{aligned} \tag{A.4}$$

By induction on n, we easily get:

$$\Gamma_k^{(n)}(f) = \sum_{i=0}^{n} \binom{n}{i} (-1)^i P_{k-n,k-i}\left[(P_{k-i,k}(f))^2\right]. \tag{A.5}$$

Since $P_{k-i,k}(f_k) = f_{k-i}$ and $f_0 = p$, it comes in the particular case of $f = f_k$:

$$\Gamma_k^{(n)}(f_k) = \sum_{i=0}^{n} \binom{n}{i} (-1)^i P_{k-n,k-i}\left(f_{k-i}^2\right).$$

Since for $f \in \mathcal{F}_k$, $\Gamma_k^{(n)}(f) \in \mathcal{F}_{k-n}$, we can compute $\gamma_{k-n}(\Gamma_k^{(n)}(f))$. From (A.5) and the fact that $\gamma_{k_n} P_{k_n,k-i} = \gamma_{k-i}$, we get:

$$\gamma_{k-n}(\Gamma_k^{(n)}(f_k)) = \sum_{i=0}^{n} \binom{n}{i} (-1)^i \gamma_{k-i}(f_{k-i}^2),$$

with $\gamma_0(f) = f$. $\Gamma_k^{(k)}(f_k)$ being a constant function in $\mathcal{F}_0$ equals $\gamma_0(\Gamma_k^{(k)}(f_k))$ and then the previous identity leads to:

$$\Gamma_k^{(k)}(f_k) = \sum_{i=0}^{k} \binom{k}{i} (-1)^i \gamma_{k-i}(f_{k-i}^2).$$

The classical inversion formula which states the equivalence between the two following identities:

$$u_k = \sum_{j=0}^{k} (-1)^j \binom{k}{j} v_j \quad \text{and} \quad v_k = \sum_{j=0}^{k} (-1)^j \binom{k}{j} u_j$$

yields that:

$$\gamma_k(f_k^2) = \sum_{j=0}^{k} \binom{k}{j} \Gamma_j^{(j)}(f_j),$$

which means that $\gamma_k(f_k^2)$ can be written as the sum of terms involving the operators Γ_k and their iterates. In particular, we get the following identity:

$$\gamma_k\left(\Gamma_{k+1}(f_{k+1})\right) = \gamma_{k+1}(f_{k+1}^2) - \gamma_k(f_k^2) = \sum_{j=0}^{k} \binom{k}{j} \Gamma_{j+1}^{(j+1)}(f_{j+1}). \tag{A.6}$$

In fact, this identity comes from a more general relation, where first we make a change of parametrization in (A.3) in order to get the following relation (valid for any function $f \in \mathcal{F}_{k+n}$):

$$P_{k+1}\left(\Gamma_{k+n}^{(n-1)}(f)\right) = \Gamma_{k+n}^{(n)}(f) + \Gamma_{k+n-1}^{(n-1)}(P_{k+n}(f)).$$

By a descendant induction on p, one gets for any $f \in \mathcal{F}_{k+1}$ and $0 \leqslant p \leqslant k$:

$$P_{p,k}\left(\Gamma_{k+1}(f)\right) = \sum_{j=p}^{k} \binom{k-p}{j-p} \Gamma_{j+1}^{(j+1-p)}\left(P_{j+1,k+1}(f)\right).$$

If $f = f_{k+1}$, since $P_{j+1,k+1}(f_{k+1}) = f_{j+1}$, we get:

$$P_{p,k}\left(\Gamma_{k+1}(f_{k+1})\right) = \sum_{j=p}^{k} \binom{k-p}{j-p} \Gamma_{j+1}^{(j+1-p)}(f_{j+1}). \tag{A.7}$$

It suffices to set $p = 0$ to recover equation (A.6) since $P_{0,k}(g) = \gamma_1 P_{1,k}(g) = \gamma_k(g)$ for any function $g \in \mathcal{F}_k$. The use of (A.2) allows us to rewrite (A.7) in the following way:

$$P_{p,k+1}(f_{k+1}^2) - P_{p,k}(f_k^2) = \sum_{j=p}^{k} \binom{k-p}{j-p} \Gamma_{j+1}^{(j+1-p)}(f_{j+1}),$$

and by a summation on k from 0 to p, for all $0 \leqslant p \leqslant k$, we get:

$$P_{p,k+1}(f_{k+1}^2) - f_p^2 = \sum_{m=p}^{k} \sum_{j=p}^{m} \binom{m-p}{j-p} \Gamma_{j+1}^{(j+1-p)}(f_{j+1}) = \sum_{j=p}^{k} \binom{k-p+1}{k-j} \Gamma_{j+1}^{(j+1-p)}(f_{j+1}).$$

When $p = 0$, we get:

$$\gamma_{k+1}(f_{k+1}^2) - f_0^2 = \sum_{j=0}^{k} \binom{k+1}{k-j} \Gamma_{j+1}^{(j+1)}(f_{j+1}),$$

which would have been also derived directly by a telescopic sum of (A.6). The action of the measure γ_p $(p \leqslant k)$ on (A.7) leads to:

$$\gamma_k \left(\Gamma_{k+1}(f_{k+1}) \right) = \sum_{j=p+1}^{k+1} \binom{k-p}{j-p-1} \gamma_p \left(\Gamma_j^{(j-p)}(f_j) \right).$$

This formula could be exploited to split the expression (12) of the variance $\mathrm{Var}(\widehat{p}_{M+1})$ into two parts:

$$\sum_{i=0}^{l} \frac{1}{r_i} \gamma_i \left(\Gamma_{i+1}(f_{i+1}) \right) + \gamma_l(D_{l+1,M}),$$

where $D_{l+1,M}$ is a quantity which depends only on the thresholds greater than l.

Appendix B. Proofs

Proof of Proposition 3.1 By the previous notation, Equations (9) and (8), the variance of the estimator can be written as:

$$\mathrm{Var}(\widehat{p}_{M+1}) = \frac{1}{r_M^2} \mathrm{Var}(Z_{M+1}) = \frac{1}{r_M^2} \sum_{i,j} \mathrm{Cov}\left(Y_{M+1}^i, Y_{M+1}^j \right).$$

To compute the covariances in the right-hand side of the previous equation, we use the classical formula:

$$\mathrm{Cov}(Y, Z) = \mathrm{Cov}(\mathbb{E}[Y|\mathcal{F}], \mathbb{E}[Z|\mathcal{F}]) + \mathbb{E}[\mathrm{Cov}(Y, Z|\mathcal{F})], \tag{B.1}$$

where Y and Z are two r.v.s and $\mathcal{F}$ a σ-algebra and $\mathrm{Cov}(Y, Z|\mathcal{F}) \coloneqq \mathbb{E}[YZ|\mathcal{F}] - \mathbb{E}[Y|\mathcal{F}]\mathbb{E}[Z|\mathcal{F}]$. In our case, conditioning with respect to the σ-algebra generated by $\mathbf{Z}_M$ leads to, for any (i, j):

$$\mathrm{Cov}(Y_{M+1}^i, Y_{M+1}^j) = R_M^2 f_M(i) f_M(j) \mathrm{Cov}(Z_{Mi}, Z_{Mj}) + R_M f_M(i)(1 - f_M(i)) \mathbb{E}[Z_{Mi}] \delta_{ij}.$$

The last term in the right-hand side cancels for $i \neq j$ since conditional on $\mathbf{Z}_M$ the r variables Y_{M+1}^i, $i = 1, \ldots, r$ are mutually independent. Finally, introducing the covariance matrix $\Sigma_n(i, j) = \mathrm{Cov}(Z_{ni}, Z_{nj})$ and using (10), we derive the following expression:

$$\mathrm{Var}(Z_{M+1}) = R_M^2 \|f_M\|_{\Sigma_M}^2 + r_M \gamma_M (f_M(1 - f_M)),$$

where $\|\cdot\|_{\Sigma_M}$ is the norm associated with the scalar product $\langle,\rangle_{\Sigma_M}$ defined by:

$$\langle f,g\rangle_{\Sigma_M} = \sum_{ij} f(i)g(j)\Sigma_M(i,j);$$

where f and g are two functions defined on $\{1,\dots,s\}$.
To compute the scalar product $\langle f,g\rangle_{\Sigma_M}$, by induction we derive the matrix Σ_M and more generally the matrices Σ_n. The initial term Σ_1 is given by (B.1), and can be rewritten as:

$$\Sigma_1(i,j) = \begin{cases} -N\gamma_1(i)\gamma_1(j), & i \neq j \\ N\gamma_1(i)(1-\gamma_1(i)), & i = j \end{cases};$$

where we get $\langle f,g\rangle_{\Sigma_1} = N\left(\gamma_1(fg) - \gamma_1(f)\gamma_1(g)\right)$. By Equation (7), we get $\Sigma_n(l,k) = \sum_{i,j}\operatorname{Cov}(Y^i_{nl}, Y^j_{nk})$ and conditioning by $\mathbf{Z}_{n-1}$, we have for $i=j$ to consider the two terms of the right-hand side of (B.1); while for $i \neq j$, the last term cancels by conditional independence. The moment generating function of the random vector $\mathbf{Y}^i_n$, conditionally to $\mathbf{Z}_{(n-1)i}$ is given by:

$$\varphi(t_1,\dots,t_r) = \left[(1-g_{n-1}(i)) + \sum_{j=1}^{r} P_n(i,j)e^{t_j}\right]^{R_{n-1}Z_{(n-1)i}}.$$

By derivation of φ (or using the multinomial distribution), we get directly that on the one hand:

$$\mathbb{E}[Y^i_{nl}|\mathbf{Z}_{n-1}] = R_{n-1}P_n(i,l)Z_{(n-1)i},$$

while on the other hand:

$$\operatorname{Cov}(Y^i_{nl}, Y^i_{nk}|\mathbf{Z}_{n-1}) = \begin{cases} R_{n-1}P_n(i,l)(1-P_n(i,l))Z_{(n-1)i} & k = l \\ -R_{n-1}P_n(i,l)P_n(i,k)Z_{(n-1)i} & k \neq l. \end{cases}$$

By Equation (10) and $\mathbb{E}[Z_{(n-1)i}] = r_{n-2}\gamma_{n-1}(i)$, we have:

$$\operatorname{Cov}(Y^i_{nl}, Y^i_{nk}) = \begin{cases} R^2_{n-1}P_n(i,l)^2\Sigma_{n-1}(i,i) + r_{n-1}\gamma_{n-1}(i)P_n(i,l)(1-P_n(i,l)) & k = l \\ R^2_{n-1}P_n(i,l)P_n(i,k)\Sigma_{n-1}(i,i) - r_{n-1}\gamma_{n-1}(i)P_n(i,l)P_n(i,k) & k \neq l, \end{cases}$$

and for $i \neq j$, $\operatorname{Cov}(Y^i_{nl}, Y^j_{nk}) = R^2_{n-1}P_n(i,l)P_n(i,k)\Sigma_{n-1}(i,j)$, that leads to the expression of $\Sigma_n(l,k)$ after a summation on i and j. Now:

$$\begin{aligned}\langle f,g\rangle_{\Sigma_n} &= \sum_{(NR_1\ldots R_{n-1})\gamma_n(j).k,l} f(k)g(l)\Sigma_n(l,k) \\ &= R^2_{n-1}\sum_{i,j,k,l} f(k)P_n(i,k)g(l)P_n(j,k)\Sigma_{n-1}(i,j) + A \\ &= R^2_{n-1}\langle P_n(f), P_n(g)\rangle_{\Sigma_{n-1}} + A,\end{aligned}$$

where:

$$\begin{aligned}A &= r_{n-1}\sum_{i,l}\gamma_{n-1}(i)f(l)g(l)P_n(i,l) - r_{n-1}\sum_{i,l,k}\gamma_{n-1}(i)f(k)g(l)P_n(i,l)P_n(i,k) \\ &= r_{n-1}\gamma_{n-1}[P_n(fg) - P_n(f)P_n(g)] = r_{n-1}\gamma_{n-1}(\Gamma_n(f,g)).\end{aligned}$$

We are lead to the following induction relation:

$$\langle f, g\rangle_{\Sigma_n} = R_{n-1}^2 \langle P_n(f), P_n(g)\rangle_{\Sigma_{n-1}} + r_{n-1}\gamma_{n-1}(\Gamma_n(f,g)),$$

that, applied to the function f_n, yields $\|f_n\|_{\Sigma_n}^2 = R_{n-1}^2\|f_{n-1}\|_{\Sigma_{n-1}}^2 + r_{n-1}\gamma_{n-1}(\Gamma_n(f_n))$, from which we deduce:

$$\begin{aligned}\frac{\mathrm{Var}(Z_{M+1})}{r_M^2} &= \frac{1}{N}\left[\gamma_1(f_1^2) - \gamma_1^2(f_1)\right] + \sum_{i=0}^{M-2} \frac{1}{r_{M-(i+1)}}\gamma_{M-(i+1)}\left(\Gamma_{M-i}(f_{M-i})\right) + \frac{1}{r_M}\gamma_M[f_M(1-f_M)] \\ &= \frac{1}{r_0}\left[\gamma_1(f_1^2) - \gamma_1^2(f_1)\right] + \sum_{i=1}^{M-1} \frac{1}{r_i}\gamma_i\left(\Gamma_{i+1}(f_{i+1})\right) + \frac{1}{r_M}\gamma_M[f_M(1-f_M)].\end{aligned}$$

With the convention and Equation (2), we get:

$$\frac{\mathrm{Var}(\widehat{p}_{M+1})}{p^2} = \sum_{i=0}^{M} \frac{1}{r_i\gamma_i(\mathbf{1})}\frac{\mu_i\left(\Gamma_{i+1}(f_{i+1})\right)}{\mu_i^2(f_i)}. \tag{B.2}$$

Proceeding with the classical notation, valid for any probability μ:

$$\mathbb{E}_\mu(f) \coloneqq \mu(f), \quad \mathrm{Var}_\mu(f) \coloneqq \mu(f^2) - \mu^2(f),$$

and using relation (5), $\gamma_{M+1}(\mathbf{1}) = p$ and $\gamma_0(\mathbf{1}) = 1$, we get the desired result. ■

Proof of Proposition 3.2 The cost of the first step of the algorithm (particles issued from 0) is $Nc_0 = r_0\gamma_0(c_0)$ and the one of the n-th step (particles issued from ∂B_{n-1}) for $n = 2, \dots, M+1$ is:

$$\sum_{i=1}^{s} R_{n-1}Z_{(n-1)i}c_{n-1}(i).$$

Finally, Formula (10) leads to a mean total cost given by (13) since by convention $\gamma_0(c_0) = c_0$.■

Proof of Proposition 4.1 The variance of the estimator is given by:

$$\frac{\mathrm{Var}(\widehat{p}_{M+1})}{p^2} = \sum_{k=1}^{M} \frac{1}{\gamma_k(\mathbf{1})}\left(\frac{1}{r_{k-1}} - \frac{1}{r_k}\right)\frac{\mathrm{Var}_{\mu_k}(f_k)}{\mathbb{E}_{\mu_k}^2(f_k)} + \sum_{k=0}^{M}\frac{1-\mu_k(g_k)}{r_k\gamma_{k+1}(\mathbf{1})}.$$

The minimisation consists of a first step in order to cancel the terms (independent of the others):

$$\frac{\mathrm{Var}_{\mu_k}(f_k)}{\mathbb{E}_{\mu_k}^2(f_k)},$$

which leads us to take the functions f_k constant on B_k, that is to require that the success probability from $\partial B_k^{(i)}$ does not depend on i. Then we are lead to the unidimensional setting and we fix $s = 1$. γ_k and g_k are now real numbers between 0 and 1:

$$\gamma_k \equiv \gamma_k(\mathbf{1}) = \mathbb{P}(\tau_k < \infty) \quad \text{and} \quad g_k \equiv \mathbb{P}(\tau_{k+1} < \infty | \tau_k < \infty). \tag{B.3}$$

In the second step, we minimise the other term of the variance for a fixed cost. The variance and the cost can be rewritten in the following way:

$$\frac{\mathrm{Var}(\widehat{p}_{M+1})}{p^2} = \sum_{k=0}^{M} \frac{1-\mu_k(g_k)}{r_k\gamma_{k+1}} \quad \text{and} \quad C = \sum_{n=0}^{M} r_n\gamma_n c_n = Nc_0 + \sum_{n=1}^{M} r_n\gamma_n c_n.$$

From (B.3), we are lead to the optimisation problem with $s = 1$ of Lagnoux-Renaudie (2008). ■

Proof Proposition 5.1 To compute the variance of the estimator $\widehat{p}_M^{(-k)}$ in the new setting, that is, without the k-th threshold, we use formula (B.2). In particular, the $(k-1)$-th first terms are unchanged, while, as we need to transport the function f_{k+1} from ∂B_{k+1} on ∂B_{k-1}, the k-th term becomes:

$$\frac{1}{\lambda_{k-1}r_k}\gamma_{k-1}(\tilde{\Gamma}_k(f_{k+1})) = \frac{1}{\lambda_{k-1}r_k}\gamma_k(\Gamma_{k+1}(f_{k+1})) + \gamma_{k-1}(\Gamma_k(f_k)),$$

where $\tilde{\Gamma}_k(f_{k+1}) = P_{k-1,k+1}(f_{k+1}^2) - [P_{k-1,k+1}(f_{k+1})]^2$. Finally, the last terms are not modified, except for the replication numbers.

Defining $\Lambda_p = \prod_{j=k-1}^{p} \lambda_j$, the variance $\mathrm{Var}\left(\widehat{p}_M^{(-k)}\right)$ of the new estimator can be expressed as:

$$\sum_{j=0}^{k-2} \frac{1}{r_j}\gamma_j\left(\Gamma_{j+1}(f_{j+1})\right) + \frac{1}{\Lambda_{k-1}r_k}\gamma_{k-1}\left(\Gamma_k(f_k)\right) + \sum_{j=k}^{M} \frac{1}{\Lambda_{j-1}r_j}\gamma_j\left(\Gamma_{j+1}(f_{j+1})\right),$$

which leads to the result.

In our context, all the c_k's are equal to 1, so the value of Λ_{k-1} given by (14) becomes:

$$\Lambda_{k-1} = \frac{1}{R_k} + \frac{\gamma_k(\mathbf{1})}{\gamma_{k-1}(\mathbf{1})} = \frac{1}{R_k} + g_{k-1}.$$

The variance is now given by:

$$\mathrm{Var}\left(\widehat{p}_{M+1}\right) = \mathrm{Var}\left(\widehat{p}_M^{(-k)}\right) + \frac{p^2}{r_{k-1}\gamma_k(\mathbf{1})}(1-g_{k-1}) + \frac{p^2}{r_{k-1}\gamma_{k+1}(\mathbf{1})}S(R_k), \tag{B.4}$$

where $S(R_k) = \frac{1}{R_k}(1-g_k) - \frac{(1-\beta)}{1+R_k g_{k-1}}$ whose minimum is achieved at:

$$R_k^* = \left(\frac{1-g_k}{1-g_{k-1}}\right)\left(1+\sqrt{\frac{1-\beta}{g_{k-1}-\beta}}\right) \quad \text{and} \quad S(R_k^*) = -(g_{k-1}-\beta)\left[\sqrt{\frac{1-\beta}{g_{k-1}-\beta}}-1\right]^2.$$

The corrective term in formula (B.4) of the variance rewrites, up to a positive multiplicative coefficient, as:

$$4g_{k-1}^2(1-\beta) - (g_{k-1}-\beta)(1+g_{k-1})^2.$$

The sign of the previous expression is the same of the polynomial $R(x) = x^2 - x(1-3\beta) + \beta$ at $x = g_{k-1}$. So:

- when $\beta > 1/9$, $R(g_{k-1})$ is strictly positive;
- when $\beta = 1/9$, $R(g_{k-1}) = (g_{k-1} - 1/3)^2$ is positive and cancels at $1/3$; and
- when $\beta < 1/9$, $R(g_{k-1})$ is minimum at $g_{k-1}^* = (1-3\beta)/2$ and $R(g_{k-1}^*) = \frac{1}{4}(1-\beta)(9\beta-1) < 0$. This minimum decreases with β from 0 (for $\beta = 1/9$) to $-1/4$ (for $\beta = 0$).

The result now becomes obvious. ■

Appendix C. Finding the Conformal Map of Section 6

The goal is to determine a conformal map φ from a disk B to a domain $\Omega := \varphi(B)$ such that the image of the occupancy measure $\varphi_* m$ on ∂B is the uniform measure on $\partial\Omega$ and $\varphi(\partial B) = \partial\Omega$. First, we restrict B to the unit disk. Since, for any Borel set $E \in \partial\Omega$:

$$\varphi_* m(E) = m(\varphi^{-1}(E)) \overset{\text{def}}{=} \int_{\partial B} \mathbb{1}_{\varphi^{-1}(E)} h(\xi) d\xi = \int_{\partial\Omega} \mathbb{1}_E(\omega) h(\varphi^{-1}(\omega)) \frac{d\omega}{|\varphi'(\varphi^{-1}(\omega))|},$$

and we want $\varphi_* m(E) = \frac{1}{|\partial\Omega|} \int_{\partial\Omega} \mathbb{1}_E(\omega) d\omega$, the conformal map φ has to satisfy $|\varphi'(\xi)| = h(\xi)|\partial\Omega|, \quad \forall \xi \in \partial B$.

Taking $|\varphi'(\xi)| = h(\xi)$ induces $|\partial\Omega| = 1$. Since φ is a conformal map, φ' is holomorphic on B and not null and $\log|\varphi'| = \log h$ is thus harmonic on B. Then we follow the procedure described in Weber and Gotsman (2010).

1. Since we work on the unit disk, we solve the Dirichlet problem and find its harmonic conjugate function concomitantly using the Schwarz integral formula (Remmert, 1991, Chap VII, §2) that allows us to recover a holomorphic function, up to an imaginary constant, from the boundary values of its real part:

$$\phi(z) = \int_0^{2\pi} \log h(e^{i\theta}) \frac{e^{i\theta} + z}{e^{i\theta} - z} \frac{d\theta}{2\pi} + ig(0), \ |z| < 1;$$

2. Now we consider e^{ϕ} which is holomorphic on B. Since B is a simply connected set and taking the Cauchy integral, there exists a holomorphic function Φ on B, such that:

$$\Phi(z) = \int_{[0,z]} e^{\phi(\omega)} d\omega,$$

   where $[0, z]$ is the segment that links 0 and z. Since e^{ϕ} never cancels, Φ is a conformal map. Thus, we define $\varphi = \Phi$.

In the case of a disk B of radius l, we take $|\varphi'(\xi)| = h(\xi)2\pi l$ instead of $|\varphi'(\xi)| = h(\xi)$ to get a boundary of length $2\pi l = |\partial\Omega|$.

Bibliography

Ahlfors, L. V., 2006. Lectures on quasiconformal mappings, 2nd Edition. Vol. 38 of University Lecture Series. American Mathematical Society, Providence, RI, with supplemental chapters by C. J. Earle, I. Kra, M. Shishikura and J. H. Hubbard.
URL https://dx-doi-org.docadis.ups-tlse.fr/10.1090/ulect/038

Aldous, D., 1989. Probability approximations via the Poisson clumping heuristic. Vol. 77 of Applied Mathematical Sciences. Springer-Verlag, New York.

Aldous, D., Vazirani, U., 1994. "go with the winners" algorithms. IEEE Symposium on Fundations of Computer Science 7, 492–501.

Antonini, R., 1991. Sur le comportement asymptotique du processus de Ornstein-Uhlenbeck multidimensionnel. Ann. Sci. Univ. Clermont-Ferrand II Probab. Appl. 9, 33–44.

Au, S., Beck, J., 1999. A new adaptive importance sampling scheme for reliability calculations. Structural Safety 21 (2), 135 – 158.
URL http://www.sciencedirect.com/science/article/pii/S0167473099000144

Au, S., Papadimitriou, C., Beck, J., 1999. Reliability of uncertain dynamical systems with multiple design points. Structural Safety 21 (2), 113 – 133.
URL http://www.sciencedirect.com/science/article/pii/S0167473099000090

Au, S.-K., Beck, J. L., 2001. Estimation of small failure probabilities in high dimensions by subset simulation. Probabilistic Engineering Mechanics 16 (4), 263–277.

Cérou, F., Del Moral, P., Guyader, A., 2011. A nonasymptotic theorem for unnormalized Feynman-Kac particle models. Ann. Inst. Henri Poincaré Probab. Stat. 47 (3), 629–649.

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

Dean, T., Dupuis, P., 2009. Splitting for rare event simulation: A large deviation approach to design and analysis. Stochastic Processes and their Applications 119 (2), 562 – 587.

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

Del Moral, P., Garnier, J., 11 2005. Genealogical particle analysis of rare events. Ann. Appl. Probab. 15 (4), 2496–2534.

Diaconis, P., Holmes, S., 1995. 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). Vol. 72 of IMA Vol. Math. Appl. Springer, New York, pp. 43–56.

Doucet, A., de Freitas, N., Gordon, N., 2001. An introduction to sequential Monte Carlo methods. In: Sequential Monte Carlo methods in practice. Stat. Eng. Inf. Sci. Springer, New York, pp. 3–14.

Garvels, M., 2000. The splitting method in rare event simulation. ProQuest LLC, Ann Arbor, MI, thesis (Dr.)–Universiteit Twente (The Netherlands).

Glasserman, P., Heidelberger, P., Shahabuddin, P., Zajic, T., 1996. Splitting for rare event simulation: analysis of simple cases. In: Proceedings of the 28th conference on Winter simulation. ACM Press, pp. 302–308.

Glasserman, P., Heidelberger, P., Shahabuddin, P., Zajic, T., 1998. A large deviations perspective on the efficiency of multilevel splitting. IEEE Trans. Automat. Control 43 (12), 1666–1679.

Glasserman, P., Heidelberger, P., Shahabuddin, P., Zajic, T., 1999. Multilevel splitting for estimating rare event probabilities. Oper. Res. 47 (4), 585–600.

Heinonen, J., 2006. What is . . . a quasiconformal mapping? Notices Amer. Math. Soc. 53 (11), 1334–1335.

Jerrum, M., Sinclair, A., 1997. The Markov chain Monte Carlo method: an approach to approximate counting and integration. Approximation Algorithms for NP-hard Problems. Dorit Hochbaum.

Kiureghian, A. D., Dakessian, T., 1998. Multiple design points in first and second-order reliability. Structural Safety 20 (1), 37 – 49.
URL http://www.sciencedirect.com/science/article/pii/S016747309700026X

Lagnoux, A., 2006. Rare event simulation. Probab. Engrg. Inform. Sci. 20 (1), 45–66.

Lagnoux-Renaudie, A., 2008. Effective branching splitting method under cost constraint. Stochastic Process. Appl. 118 (10), 1820–1851.

L'Ecuyer, P., Le Gland, F., Lezaud, P., Tuffin, B., 2009. Splitting techiques. In: Rare event simulation using Monte Carlo methods. Wiley, Chichester, pp. 39–61.

Miretskiy, D., Scheinhardt, W., Mandjes, M., 2009. Proceedings of the 6th st. Petersburg workshop on simulation, St. Petersburg, Russia. St. Petersburg State University, St. Petersburg, Russia, pp. 909–914.

Morio, J., Balesdent, M., Jacquemart, D., Vergé, C., 2014. A survey of rare event simulation methods for static input/output models. Simulation Modelling Practice and Theory 49 (0), 287 – 304.

Patane, G., Li, X. S., Gu, D. X., 2014. An introduction to ricci flow and volumetric approximation with applications to shape modeling. In: SIGGRAPH Asia 2014 Courses. SA '14. ACM, New York, NY, USA, pp. 4:1–4:118.
URL http://doi.acm.org/10.1145/2659467.2659469

Remmert, R., 1991. Theory of complex functions. Vol. 122 of Graduate Texts in Mathematics. Springer-Verlag, New York, translated from the second German edition by Robert B. Burckel, Readings in Mathematics.

Sadowsky, J., 1996. On Monte Carlo estimation of large deviations probabilities. Ann. Appl. Probab. 6 (2), 399–422.

Schuëller, G., Pradlwarter, H., Pandey, M., 1993. Methods for reliability assessment of nonlinear systems under stochastic dynamic loading - a review. EURODYN'93, Balkema, pp. 751–9.

Villén-Altamirano, M., Villén-Altamirano, J., 1991. Restart: a method for accelerating rare event simulations. North-Holland, pp. 71–76.

Villén-Altamirano, M., Villén-Altamirano, J., 1997. Restart: An efficient and general method for fast simulation of rare event. *Technical report 7*.

Weber, O., Gotsman, C., Jul. 2010. Controllable conformal maps for shape deformation and interpolation. ACM Trans. Graph. 29 (4), 78:1–78:11.

Wolfram Research, I., 2015. Mathematica, Version 10.