A combined importance splitting and sampling algorithm for rare event estimation

Damien Jacquemart-Tomi, Jérôme Morio, François Le Gland

Proceedings of the 2013 Winter Simulation Conference
R. Pasupathy, S.-H. Kim, A. Tolk, R. Hill, and M. E. Kuhl, eds.

A COMBINED IMPORTANCE SPLITTING AND SAMPLING ALGORITHM FOR RARE EVENT ESTIMATION

Damien Jacquemart-Tomi
Jérôme Morio

ONERA, the French Aerospace Lab
F-91761 Palaiseau, FRANCE

François Le Gland

INRIA
Campus de Beaulieu
35042 Rennes, FRANCE

ABSTRACT

We propose some methodological basis for an improvement to the splitting method for a Markov process that evolves over a deterministic time horizon. Our algorithm is based on a decomposition of the selection functions that gives more importance to some well-chosen trajectories, typically those trajectories that manage to move earlier than others towards the critical region. Central limit theorem is established and numerical experiments are provided.

1 INTRODUCTION

Rare event estimation is of great interest in many scientific and industrial areas such as air traffic management, telecommunication, nuclear engineering, climatology, *etc.* Crude Monte Carlo simulations are no longer efficient for low probabilities and specific techniques dedicated to rare events are required. When the event of interest is modelled as the first hitting time for Markov chain or process, importance sampling (Juneja and Shahabuddin 2006), weighted importance resampling (Del Moral and Garnier 2005) and importance splitting (L'Ecuyer, Demers, and Tuffin 2006) are the most valuable methods to deal with this problem.

This paper focuses on the importance splitting method. The idea is to decompose the sought probability in a product of conditional probabilities that can be estimated accurately with a reasonable computation time. Numerous variants have been worked out and presentations of the summary of the methods can be found in (L'Écuyer, Demers, and Tuffin 2007), (L'Écuyer et al. 2009). Splitting methods have notably been compared in (L'Ecuyer, Demers, and Tuffin 2006). Latest developments, linked with genetic algorithms, can be found in (Cérou et al. 2005), (Cérou et al. 2006), where rigorous proofs of convergence are given. In addition, approximations of the stochastic process law in the rare event regime are obtained. This latest variant will be used in this article. It is organised as follows. First, standard splitting algorithm is recalled, and a multidimensional adaptive algorithm is given. We propose then a way to improve splitting algorithm, by giving more importance to some selected particles. A central limit theorem with regard to this new method is established. Finally, numerical experiments on a toy case are provided.

2 STANDARD SPLITTING ALGORITHM

Splitting methods are of great interest when one has to work with stochastic processes that evolve in continuous time. The framework is the following. Given a probability space $(\Omega, \mathscr{F}, \mathbb{P})$ and a measurable metric space E, let $X : (\omega, t) \in \Omega \times [0, +\infty) \to X_t(\omega) \in E$ be a Markov process with continuous trajectories, or at least right continuous with left hand limits trajectories, let B be a measurable subset of the state space E and T_B the first hit time of B, given by

$$T_B = \inf\{t \geq 0, X_t \in B\}.$$

Given T a deterministic time, or a random stopping time with finite expectation, the event

$$\{T_B \leq T\} = \{X_t \in B, \text{ for some } t \leq T\},$$

is supposed to be very rare. Its probability is typically lower than 10^{-4}. In this context, splitting methods give efficient numerical approximations to the rare event probability $\mathbb{P}(T_B \leq T)$, and to the law of the process in the rare regime $\mathbb{E}(f(X_t, 0 \leq t \leq T_B)|T \leq T_B)$. The principle of splitting is to consider a sequence of decreasing supersets of $B \times [0,\infty)$:

$$E \times [0,\infty) \supset A_1 \supset \cdots \supset A_{m-1} \supset A_m = B \times [0,\infty), \tag{1}$$

and to estimate each probability that the process $t \mapsto (X_t,t)$ starting from A_{k-1} reaches A_k before time T.

Let us define $T_k = \inf\{t \geq 0 : (X_t,t) \in A_k\}$ for $k = 1,\ldots,m$. It is worth noting that the existence of left-hand limits ensures that $(X_{T_k},T_k) \in \overline{A}_k$ where $\overline{A}_k$ is the closure of A_k, hence $(X_{T_k},T_k) \in A_k$ provided A_k is a closed subset. From now on, it is assumed that the supersets introduced in (1) are all closed subsets. A Bayes formula gives the following product decomposition equation

$$\mathbb{P}(T_B \leq T) = \prod_{k=1}^{m} p_k \quad \text{where} \quad p_1 = \mathbb{P}(T_1 \leq T) \quad \text{and} \quad p_k = \mathbb{P}(T_k \leq T \mid T_{k-1} \leq T) \text{ for } k = 2,\ldots,m.$$

The intermediate subsets have to be chosen such that the probabilities p_k are large enough to be estimated accurately with Monte Carlo. In practice, in a first stage, N samples of Markov process (X_t,t) are generated until time $T_1 \wedge T = \min(T_1,T)$. If I_1 is the number of trajectories which have reached A_1, then $p_1 = \mathbb{P}(T_1 \leq T)$ is estimated by I_1/N. For stage $k \geq 1$, N starting points are randomly and uniformly chosen amongst the I_k crossing points between the subset A_k and the previously sampled trajectories. N paths of the process (X_t,t) are sampled from these crossing points according to the Markov dynamic of the process until time $T_{k+1} \wedge T$. If I_{k+1} is the number of trajectories that have reached A_{k+1}, then $p_{k+1} = \mathbb{P}(T_{k+1} \leq T \mid T_k \leq T)$ is estimated by I_{k+1}/N.

2.1 Feynman-Kac Interpretation of the Algorithm

We will now give a Feynman-Kac interpretation of the algorithm, slightly different from the one given in (Cérou et al. 2006) for the sake of comprehension of the sequel. To model the algorithm, it is convenient to consider the $E \times \mathbb{R}_+$ valued random variable Z_k

$$Z_k = (X_{T_k \wedge T}, T_k \wedge T).$$

The sequence Z_k is a Markov chain with initial distribution $\eta_0 \times \delta_0$, where η_0 is the distribution of X_0 and δ_0 is the Dirac mass distribution at 0, and with transition kernels

$$Q_k(x,t,\mathrm{d}x',\mathrm{d}t') = \mathbb{P}(X_{T_k \wedge T} \in \mathrm{d}x', T_k \wedge T \in \mathrm{d}t' | X_{T_{k-1} \wedge T} = x, T_{k-1} \wedge T = t).$$

Considering next the selection functions $g_k(x,t) = 1_{\{(x,t) \in A_k\}}$, it is possible to give an interpretation of the rare event probability in terms of the Feynman-Kac distribution

$$\langle \gamma_k, f \rangle = \mathbb{E}[f(Z_k) \prod_{p=1}^{k} g_k(Z_k)] = \mathbb{E}[f(X_{T_k},T_k) 1_{\{T_k \leq T\}}] \quad \text{and} \quad \langle \mu_k, f \rangle = \frac{\langle \gamma_k, f \rangle}{\langle \gamma_k, 1 \rangle}.$$

The nonnegative distributions γ_k satisfy the recurrent relation $\gamma_k = g_k(\gamma_{k-1} Q_k) = \gamma_{k-1} R_k$, where the nonnegative kernels R_k are defined by $R_k(z,\mathrm{d}z') = g_k(z') Q_k(z,\mathrm{d}z')$. One has thus $\langle \gamma_k, 1 \rangle = \mathbb{P}(T_k \leq T)$ and

$\mu_k(\mathrm{d}x,\mathrm{d}t) = \mathbb{P}(X_{T_k} \in \mathrm{d}x, T_k \in \mathrm{d}t | T_k \leq T)$ is then the entrance distribution in the subset A_k. In the same way, defining

$$\langle \eta_k, f \rangle = \mathbb{E}[f(Z_k)|T_{k-1} \leq T],$$

one obtains $\langle \eta_k, g_k \rangle = \mathbb{P}(T_k \leq T | T_{k-1} \leq T) = p_k$ and the following recursive equation $\langle \gamma_k, f \rangle = \langle \eta_k, g_k f \rangle \langle \gamma_{k-1}, 1 \rangle$. Now, the following main formula can be derived

$$\mathbb{P}(T_B \leq T) = \langle \gamma_m, 1 \rangle = \prod_{k=1}^{m} \langle \eta_k, g_k \rangle.$$

2.2 The Importance Function

The supersets $A_k \supset B \times [0,\infty)$ are often defined by threshold exceedance of a real–valued function Φ, which is called *importance function* in the splitting algorithm, *i.e.*

$$A_k = \{(x,t) \in E \times [0,\infty) : \Phi(x,t) \geq S_k\} \quad \text{and} \quad T_k = \inf\{t \geq 0, \Phi(X_t,t) \geq S_k\}, \tag{2}$$

so that $\{T_k \leq T\} = \{\Phi(X_t,t) \geq S_k, \text{ for some } 0 \leq t \leq T\} = \{\max_{0 \leq t \leq T} \Phi(X_t,t) \geq S_k\}$, for all $k = 1,\ldots,m$, with a suitable sequence of real numbers $S_1 < S_2 < \cdots < S_m = S$. In particular for the ultimate threshold $S = S_m$, one should have $T_B = \inf\{t \geq 0 : \Phi(X_t,t) \geq S\} = \inf\{t \geq 0 : X_t \in B\}$, which means that some *compatibility* condition should hold for the importance function Φ and for the threshold S to make sure that the two definitions of the event $\{T_B \leq T\}$ are consistent. Even though the optimal importance function Φ is time–dependent, it is often the case that a simpler but sub–optimal time–independent importance function Φ is used, which is already available and intrinsically given by the problem.

It is well known that the choice of a good importance function is crucial to get reliable estimation of the probability $\mathbb{P}(T_B \leq T)$ (Glasserman et al. 1998). (Cérou et al. 2006) shows that the optimal subsets A_k are obtained when the probability of reaching the critical set $B \times [0,\infty)$ starting from any possible hitting position and time $(X_{T_k}, T_k) = (x,s) \in A_k$ should not depend on (x,s). Defining optimal subsets would thus lead to know $u_B(x,s) = \mathbb{P}(X_t \in B, \text{ for some } s \leq t \leq T | X_s = x)$ for every $(x,s) \in E \times [0,\infty)$. Unfortunately, this choice is clearly unrealistic for most practical problems, since knowing $u_B(x,s)$ implies the knowledge of $\mathbb{P}(T_B \leq T)$. The paper (Garvels, Van Ommeren, and Kroese 2002) links the optimal subsets to an optimal importance function and provides methods to estimate it, but the framework is restrained to discrete state space Markov process. To determine an importance function, optimal or not, is of great interest since it enables an adaptive choice of the intermediate subsets, as explained in the next section.

2.3 Adaptive Choice of the Thresholds

When the supersets $A_k \supset B \times [0,\infty)$ are characterised as in equation (2), the thresholds S_k can easily be adaptively chosen. An adaptive choice of the thresholds was first given in (Garvels 2000), then convergence proof for the dimension one is given in (Cérou and Guyader 2007), and a multidimensional algorithm is evoked in (Cérou et al. 2006). In this article we propose another method, used in air traffic management context in (Jacquemart and Morio 2013), which enables easy implementation of the splitting algorithm in the multidimensional and non-homogeneous case. It is decomposed in two stages at each iteration $k = 1,\ldots,m$. The first stage enables to estimate intermediate thresholds and the second one determines the starting points at level A_k. For $k \geq 1$, assume that one knows the threshold S_k and an approximation of the entrance distribution μ_k, denoted by $\hat{\mu}_k$. An adaptive method to estimate the threshold S_{k+1} is to consider quantile estimation. One can indeed sample N new Markov processes $t \mapsto (X_t^i, t)$, $i = 1,\ldots,N$, starting from the entrance distribution at level A_k, namely $\hat{\mu}_k$, and determine the maxima of $\Phi(X_t^i,t)$ before final time T. The threshold S_{k+1} is defined as the $(1-p)$-quantile of these maxima, and one has thus $\mathbb{P}(T_{k+1} \leq T | T_k \leq T) \approx p$. To determine an approximation of the entrance distribution μ_{k+1}, a new set of N' trajectories is sampled from the entrance distribution $\hat{\mu}_k$ and until time T or until the first time they hit

the threshold S_{k+1}. The N_k crossing points $e_i = (X^i_{T^i_k \wedge T}, T^i_k \wedge T)$ for which $T^i_k \leq T$ define an approximation of μ_{k+1}:

$$\hat{\mu}_{k+1} \approx \frac{1}{N_k} \sum_{i=1}^{N_k} \delta_{e_i}.$$

An approximation of the conditional probability $\mathbb{P}(T_{k+1} \leq T | T_k \leq T)$ is given by N_k/N'. The algorithm has converged when $S_k \geq S$. Amongst the last sample, a proportion r of the trajectories reaches the final threshold S and one has the following estimation of the rare event probability, where m is the number of created thresholds

$$\mathbb{P}(T_B \leq T) \approx r \times \prod_{k=1}^{m} \frac{N_k}{N'}.$$

One has thus fully determined adaptively with this algorithm

- the intermediate subsets A_k, implicitly defined with the importance function Φ,
- approximations of the entrance distributions μ_k,
- an estimation of the rare event probability.

In the next section, we propose a way to improve the efficiency of the splitting algorithm when intermediate thresholds $S_1 < \ldots < S_m$ are given. They can for instance be determined by the previous procedure.

3 A COMBINED IMPORTANCE SPLITTING AND SAMPLING ALGORITHM

In the scientific literature, several methods have been worked out in order to combine in different ways the importance splitting algorithms and other statistical methods. Most of them act on the dynamic of the process. See (L'Écuyer, Demers, and Tuffin 2007) for a survey. In this article, the idea is to use different selection functions, and not to use auxiliary transitions kernels. Defining the subset A_k with a non-optimal importance function Φ is most of the time the best to do, and the adaptive algorithm given in section 2.3 is a very simple way to proceed. We assume to be in the case when one has to work with a deterministic horizon time T. Intuitively, trajectories that have reached the level A_k early are more likely to reach the rare set B than trajectories that have reached the level A_k later on. We propose thus to give more importance to them, selecting them mostly than the others. After that, a weighting step is required not to bias the estimated probability.

The following section is organised as follows. Firstly, an interpretation of the combined importance splitting and sampling (I2S) algorithm in terms of Feynman-Kac distributions is given. A pathwise point of view is required but is only a trick that naturally disappears. Then, the proposed algorithm is rewritten in term of the original continuous time Markov process $\{X_t\,, t \in [0,\infty)\}$. Finally, a central limit theorem on the probability estimation is given.

3.1 Interpretation of the Algorithm in Term of Feynman-Kac Distributions

The key idea is to decompose the potential function as $g_k = g_k^{\text{imp}} g_k^{\text{red}}$, where the function g_k^{red} is used to resample the successful trajectories, and where the function g_k^{imp} is used to measure the importance of trajectories with regards to the estimations, in other words to weight trajectories. We propose the following factorization

$$g_k^{\text{imp}}(x,t) = \frac{1}{a_k(t)} \quad \text{and} \quad g_k^{\text{red}}(x,t) = a_k(t) 1_{\{\Phi(x,t) \geq S_k\}}, \tag{3}$$

so that $g_k^{\text{red}} g_k^{\text{imp}} = g_k$, with a function $t \mapsto a_k(t)$ that should be nonincreasing, in order to select preferably those trajectories that reach the level A_k early. The unnormalised Feynman-Kac distribution γ_k can be

rewritten in the following way

$$\langle \gamma_k, f \rangle = \mathbb{E}[f(Z_k) \prod_{p=1}^{k} g_p(Z_p)] = \mathbb{E}[f(Z_k) \prod_{p=1}^{k} g_p^{\text{imp}}(Z_p) \prod_{p=1}^{k} g_p^{\text{red}}(Z_p)].$$

One can decide to group the potentials g_p^{imp} with the test function f. To introduce the representation of the model in term of Feynman-Kac distributions, it is convenient to adopt a path point of view. Let Y_k denote the historical process of the Markov chain Z_k:

$$Y_k = Z_{0:k} = (Z_0, Z_1, \ldots, Z_k).$$

The Markov chain Y_k is characterized by the probability transitions:

$$Q_k^{\bullet}(y_{k-1}, \mathrm{d}y'_k) = Q_k^{\bullet}(z_0, \ldots, z_{k-1}, \mathrm{d}z'_0, \ldots, \mathrm{d}z'_k) = \delta_{(z_0,\ldots,z_{k-1})}(\mathrm{d}z'_0, \ldots, \mathrm{d}z'_{k-1}) Q_k(z_{k-1}, \mathrm{d}z'_k)$$

where $Q_k(z_{k-1}, \mathrm{d}z'_k) = \mathbb{P}(Z_k \in \mathrm{d}z'_k | Z_{k-1} = z_{k-1})$ are the transition kernels of the Markov chain Z_k. The following normalized Feynman-Kac distributions on the path space can be defined

$$\langle \gamma_k^{\bullet}, f \rangle = \mathbb{E}[f(Y_k) \prod_{p=1}^{k} g_p^{\bullet}(Y_p)] \quad \text{and} \quad \langle \mu_k^{\bullet}, f \rangle = \frac{\langle \gamma_k^{\bullet}, f \rangle}{\langle \gamma_k^{\bullet}, 1 \rangle}$$

with the notation $g_p^{\bullet}(y_p) = g_p^{\bullet}(z_0, \ldots, z_p) = g_p^{\text{red}}(z_p)$. The unnormalized distributions satisfy the recursive equation

$$\gamma_k^{\bullet} = g_k^{\bullet}(\gamma_{k-1}^{\bullet} Q_k^{\bullet}) = g_k^{\bullet}(\mu_{k-1}^{\bullet} Q_k^{\bullet}) \langle \gamma_k^{\bullet}, 1 \rangle. \tag{4}$$

To connect the two distributions γ_k and $\gamma_k^{\bullet}$, we use the function $T_k^{\bullet} f$ defined on the path space by

$$T_k^{\bullet} f(y_k) = T_k^{\bullet} f(z_0, \ldots, z_k) = f(z_k) \prod_{p=1}^{k} g_p^{\text{imp}}(z_k),$$

so that $\langle \gamma_k, f \rangle = \mathbb{E}[f(Z_k) \prod_{p=1}^{k} g_p(Z_p)] = \mathbb{E}[T_k^{\bullet} f(Y_k) \prod_{p=1}^{k} g_p^{\bullet}(Y_p)] = \langle \gamma_k^{\bullet}, T_k^{\bullet} f \rangle$. In particular, with $f \equiv 1$, one obtains $\langle \gamma_k^{\bullet}, T_k^{\bullet} 1 \rangle = \langle \gamma_k, 1 \rangle = \mathbb{P}(T_k \leq T)$. However, the normalizing constant $\langle \gamma_k^{\bullet}, 1 \rangle$ does not seem to have any useful probabilistic interpretation.

3.2 Interacting Path-Particle Interpretation

We are interested in approximations of the normalized distribution $\mu_{k-1}^{\bullet}$ with the following form

$$\mu_{k-1}^{\bullet,N} = \sum_{i=1}^{N} w_{k-1}^{i} \delta_{\xi_{k-1}^{\bullet,i}} \text{ , with } \sum_{j=1}^{N} w_{k-1}^{j} = 1$$

where, for all $i = 1, \ldots, N$ the particle $\xi_{k-1}^{\bullet,i}$ is a random trajectory $\xi_{k-1}^{\bullet,i} = (\xi_{0,k-1}^{i}, \ldots, \xi_{k-1,k-1}^{i})$. Furthermore, motivated by equation (4) one can write

$$\mu_{k-1}^{\bullet,N} Q_k^{\bullet}(\mathrm{d}y'_k) = \sum_{i=1}^{N} w_{k-1}^{i} \, Q_k^{\bullet}(\xi_{k-1}^{\bullet,i}, \mathrm{d}y'_k), \tag{5}$$

which yields to the particle approximation $\eta_k^{\bullet,N} = \dfrac{1}{N} \sum_{i=1}^{N} \delta_{\xi_k^{\bullet,i}}$ of the distribution $\eta_k^{\bullet} = \mu_{k-1}^{\bullet} Q_k^{\bullet}$, where independently for all $i = 1, \ldots, N$, the trajectory $\xi_k^{\bullet,i}$ is sampled from the finite mixture distribution (5). In practice, the transition from $\mu_{k-1}^{\bullet,N}$ to $\eta_k^{\bullet,N}$ is obtained as follows. Independently for all $i = 1, \ldots, N$

1. a path $\widehat{\xi}^{\bullet,i}_{k-1} = (\widehat{\xi}^i_{0,k-1},\ldots,\widehat{\xi}^i_{0,k-1})$ is selected amongst the current population $(\xi^{\bullet,1}_{k-1},\ldots,\xi^{\bullet,N}_{k-1})$, according to their respective weights $(w^1_{k-1},\ldots,w^N_{k-1})$.
2. the random path $\xi^{\bullet,i}_k = (\xi^i_{0,k},\ldots,\xi^i_{k,k})$ is sampled from the distribution $Q^\bullet_k(\widehat{\xi}^{\bullet,i}_{k-1},\mathrm{d}y'_k)$: in other words, we set $\xi^i_{p,k} = \widehat{\xi}^i_{p,k-1}$ for all $p = 0,\ldots,(k-1)$, and the random variable $\xi^i_{k,k}$ is sampled from the distribution $Q_k(\xi^i_{k-1,k},\mathrm{d}z'_k)$.

With the recursive equation (4), one obtains the following particle approximation of the unnormalized distribution $\gamma^\bullet_k$:

$$\gamma^{\bullet,N}_k = g^\bullet_k \eta^{\bullet,N}_k \langle \gamma^{\bullet,N}_{k-1}, 1\rangle = \left(\frac{1}{N}\sum_{i=1}^N g^\bullet_k(\xi^{\bullet,i}_k)\delta_{\xi^{\bullet,i}_k}\right)\langle \gamma^{\bullet,N}_{k-1}, 1\rangle.$$

From there, one can deduce several useful approximations

- the approximation of the normalization constant as a recursive equation

$$\langle \gamma^{\bullet,N}_k, 1\rangle = \left(\frac{1}{N}\sum_{i=1}^N g^\bullet_k(\xi^{\bullet,i}_k)\right)\langle \gamma^{\bullet,N}_{k-1}, 1\rangle,$$

- the particle approximation of the normalized distribution $\mu^{\bullet,N}_k$ as a recursive equation

$$\mu^{\bullet,N}_k = \frac{\gamma^{\bullet,N}_k}{\langle \gamma^{\bullet,N}_k, 1\rangle} = \sum_{i=1}^N \frac{g^\bullet_k(\xi^{\bullet,i}_k)}{\sum_{j=1}^N g^\bullet_k(\xi^{\bullet,j}_k)}\delta_{\xi^{\bullet,i}_k} = \sum_{i=1}^N w^i_k \delta_{\xi^{\bullet,i}_k},$$

which implicitly defines the weight $w^i_k = \dfrac{g^\bullet_k(\xi^{\bullet,i}_k)}{\sum_{j=1}^N g^\bullet_k(\xi^{\bullet,j}_k)}$ for all $i = 1,\ldots,N$,

- the approximation of the probability $P_k := \mathbb{P}(T_k \le T) = \langle \gamma^\bullet_k, T^\bullet_k 1\rangle$ as

$$P^N_k = \langle \gamma^{\bullet,N}_k, T^\bullet_k 1\rangle = \left(\frac{1}{N}\sum_{i=1}^N g^\bullet_k(\xi^{\bullet,i}_k) T^\bullet_k 1(\xi^{\bullet,i}_k)\right)\langle \gamma^{\bullet,N}_{k-1}, 1\rangle.$$

It is now possible to re-interpret these formulae with the potential functions g^{red}_k and g^{imp}_k in the following way. To this end, let $\xi^i_k = \xi^i_{k,k}$ denote the final state of the trajectory $\xi^{\bullet,i}_k = (\xi^i_{0,k},\ldots,\xi^i_{k,k})$. One has then $g^\bullet_k(\xi^{\bullet,i}_k) = g^{\mathrm{red}}_k(\xi^i_k)$, and if one sets $v^i_k = T^\bullet_k 1(\xi^{\bullet,i}_k)$, one remarks also that

$$v^i_k = \prod_{p=1}^k g^{\mathrm{imp}}_p(\xi^i_{p,k}) = g^{\mathrm{imp}}_k(\xi^i_{k,k})\prod_{p=1}^{k-1} g^{\mathrm{imp}}_p(\xi^i_{p,k}) = g^{\mathrm{imp}}_k(\xi^i_{k,k})\prod_{p=1}^{k-1} g^{\mathrm{imp}}_p(\widehat{\xi}^i_{p,k-1}).$$

One deduces then

$$\langle \gamma^{\bullet,N}_k, 1\rangle = \left(\frac{1}{N}\sum_{i=1}^N g^{\mathrm{red}}_k(\xi^i_k)\right)\langle \gamma^{\bullet,N}_{k-1}, 1\rangle \quad \text{and} \quad P^N_k = \left(\frac{1}{N}\sum_{i=1}^N g^{\mathrm{red}}_k(\xi^i_k) v^i_k\right)\langle \gamma^{\bullet,N}_{k-1}, 1\rangle = \frac{\sum_{i=1}^N g^{\mathrm{red}}_k(\xi^i_k)\, v^i_k}{\sum_{i=1}^N g^{\mathrm{red}}_k(\xi^i_k)}\langle \gamma^{\bullet,N}_k, 1\rangle.$$

Here finally comes the following recursive non pathwise implementation, in terms of a mixed population, where for all $i = 1,\ldots,N$ the particle $\Xi^i_k = (\xi^i_k, v^i_k)$ is a pair (position, auxiliary weight), evolving in the following way. Independently, for all $i = 1,\ldots,N$

1. a pair $\widehat{\Xi}_{k-1}^i = (\widehat{\xi}_{k-1}^i, \widehat{v}_{k-1}^i)$ is selected among the current population $(\Xi_{k-1}^1, \ldots, \Xi_{k-1}^N)$, according to their respective weights $(w_{k-1}^1, \ldots, w_{k-1}^N)$,
2. the variable ξ_k^i is sampled with the distribution $Q_k(\widehat{\xi}_{k-1}^i, \mathrm{d}z')$, and we set $v_k^i = g_k^{\mathrm{imp}}(\xi_k^i)\, \widehat{v}_{k-1}^i$,
3. the normalized weight is defined as $w_k^i \propto g_k^{\mathrm{red}}(\xi_k^i)$.

This procedure is rewritten below in algorithm 1 in terms of the original continuous time Markov process $\{X_t, t \in [0,\infty)\}$ and in the special case $g_k^{\mathrm{imp}}(x,t) = \dfrac{1}{a_k(t)}$ and $g_k^{\mathrm{red}}(x,t) = a_k(t) 1_{\{\Phi(x,t) \geq S_k\}}$.

Algorithm 1: I2S algorithm

Initialisation

Set $J_0 = \{1, \ldots, N\}$ and $\widehat{\theta}_0 = 0$.
for $i = 1, \ldots, N$ **do**
Set $T_0^i = 0$ and sample X_0^i independently from law η_0.
Set $v_0^i = 1/N$ and $w_0^i = 1/N$ the initial importance weights and resampling weights.

for $k = 1, \ldots, m$ **do**
for $i = 1, \ldots, N$ **do**
Select an index $I \in J_{k-1}$ from the weighted law $\displaystyle\sum_{j \in J_{k-1}} w_{k-1}^j \delta_j$.
Set $\widehat{X}_{k-1}^i = X_{k-1}^I$, $\widehat{T}_{k-1}^i = T_{k-1}^I$ and $\widehat{v}_{k-1}^i = v_{k-1}^I$.
Sample a path X_t^i starting from state $\widehat{X}_{k-1}^i$ at time $\widehat{T}_{k-1}^i$ until time $T_k^i \wedge T$,
where $T_k^i = \inf\{t \geq \widehat{T}_{k-1}^i : \Phi(X_t^i, t) \in A_k\}$.
Set the importance weights v_k^i and the resampling weights w_k^i as
$$v_k^i = \frac{1}{a_k(T_k^i)}\, \widehat{v}_{k-1}^i \quad \text{and} \quad w_k^i = a_k(T_k^i)\, 1_{\{T_k^i \leq T\}}.$$
Normalize the resampling weights.
Set $J_k = \{i = 1, \ldots, N, T_k^i \leq T\}$ and define
$$\widehat{\theta}_k = \left(\frac{1}{N} \sum_{i \in J_k} a_k(T_k^i) \right) \widehat{\theta}_{k-1} \quad \text{and} \quad \widehat{P}_k^{\mathrm{I2S}} = \frac{\displaystyle\sum_{i \in J_k} a_k(T_k^i)\, v_k^i}{\displaystyle\sum_{i \in J_k} a_k(T_k^i)}\, \widehat{\theta}_k$$

Estimation

$\mathbb{P}(T_B \leq T) \approx \widehat{P}_m^{\mathrm{I2S}}$

3.3 Variance Analysis

The main result of this section is the central limit theorem for the estimation of the rare event probability with the combined importance sampling and importance splitting algorithm. It is given in terms of Feynman-Kac distributions defined in (6). A more explicit expression in terms of the functions $a_k(t)$ is obtained taking into account the special form (3) of the factorization.

Theorem 1 The I2S algorithm is unbiased. Furthermore, the central limit theorem for the relative error of the estimator $\widehat{P}_m^{\mathrm{I2S}}$ given by the I2S algorithm stands as follows

$$\sqrt{N}\, \frac{\widehat{P}_m^{\mathrm{I2S}} - P_B}{P_B} \underset{N \to +\infty}{\overset{\mathrm{Law}}{\Longrightarrow}} \mathcal{N}(0, V_m),$$

where

$$V_m = \left[\frac{\langle \eta_0, (R_{1:m}1)^2 \rangle}{\langle \eta_0, R_{1:m}1 \rangle^2} - 1 \right] + \sum_{k=1}^{m} \left[\langle \gamma_{k-1}^{\mathrm{red}}, 1 \rangle \frac{\langle \gamma_k^{\mathrm{imp}}, g_k^{\mathrm{red}} (R_{k+1:m}1)^2 \rangle}{\langle \gamma_k, R_{k+1:m}1 \rangle^2} - 1 \right].$$

Here, the functions $R_{k+1:m}f$ are defined inductively by

$$R_{k:m}f = R_k(R_{k+1:m}f) \quad \text{and} \quad R_{m+1:m}f = f \quad \text{by convention,}$$

and have the probabilistic interpretation $R_{k+1:m}f(z) = \mathbb{E}[f(Z_m) \prod_{p=k+1}^{m} g_p(Z_p) | Z_k = z]$. Explicit expression of the variance can also easily be obtained in terms of the functions $a_p(t)$ and $u_B(x,t) = \mathbb{P}(T_B \leq T | X_t = x)$.

Proof. When importance sampling in added to the multilevel algorithm, the variance takes the form (Le Gland 2007):

$$V_m = \frac{\mathrm{var}(R_{1:m}1, \eta_0)}{\langle \eta_0, R_{1:m}1 \rangle^2} + \sum_{k=1}^{m} C_k + \sum_{k=1}^{m} D_k,$$

where $C_k = \dfrac{\langle \gamma_{k-1}^{\mathrm{imp}}, 1 \rangle \langle \gamma_{k-1}^{\mathrm{red}}, 1) \rangle}{\langle \gamma_{k-1}, 1 \rangle^2} \dfrac{\mathrm{var}(g_k\, R_{k+1:m}1, \mu_{k-1}^{\mathrm{imp}} Q_k)}{\langle \mu_{k-1}, R_{k:m}1 \rangle^2}$ and $D_k = \dfrac{\langle \gamma_{k-1}^{\mathrm{imp}}, 1 \rangle \langle \gamma_{k-1}^{\mathrm{red}}, 1 \rangle}{\langle \gamma_{k-1}, 1 \rangle^2} \dfrac{\langle \mu_{k-1}^{\mathrm{imp}}, R_{k:m}1 \rangle^2}{\langle \mu_{k-1}, R_{k:m}1 \rangle^2} - 1$.The distribution μ_k^{red} and μ_k^{imp} are the normalised distributions associated with γ_k^{red} and γ_k^{imp}, namely

$$\mu_k^{\mathrm{red}} = \frac{\gamma_k^{\mathrm{red}}}{\langle \gamma_k^{\mathrm{red}}, 1 \rangle} \quad \text{and} \quad \mu_k^{\mathrm{imp}} = \frac{\gamma_k^{\mathrm{imp}}}{\langle \gamma_k^{\mathrm{imp}}, 1 \rangle},$$

and following (Le Gland 2007), the unnormalised distributions γ_k^{red} and γ_k^{imp} are defined as follows

$$\langle \gamma_k^{\mathrm{red}}, f \rangle = \mathbb{E}[f(Z_k) \prod_{p=1}^{k} g_p^{\mathrm{red}}(Z_p)] \quad \text{and} \quad \langle \gamma_k^{\mathrm{imp}}, f \rangle = \mathbb{E}[f(Z_k) \prod_{p=1}^{k} \big(g_p^{\mathrm{imp}}(Z_p)\big)^2 \prod_{p=1}^{k} g_p^{\mathrm{red}}(Z_p)]. \tag{6}$$

In the original paper (Le Gland 2007), γ_k^{imp} is denoted as $\gamma_k^{\square}$, and it is rewritten here for the sake of clarity. The nonnegative distributions γ_k^{red} and γ_k^{imp} satisfy the recurrent relations

$$\gamma_k^{\mathrm{red}} = \gamma_{k-1}^{\mathrm{red}}\, R_k^{\mathrm{red}} \quad \text{and} \quad \gamma_k^{\mathrm{imp}} = \gamma_{k-1}^{\mathrm{imp}}\, R_k^{\mathrm{imp}},$$

where the nonnegative kernels R_k^{red} and R_k^{imp} are defined by

$$R_k^{\mathrm{red}}(z, \mathrm{d}z') = g_k^{\mathrm{red}}(z')\, Q_k(z, \mathrm{d}z') \quad \text{and} \quad R_k^{\mathrm{imp}}(z, \mathrm{d}z') = (g_k^{\mathrm{imp}}(z'))^2\, g_k^{\mathrm{red}}(z')\, Q_k(z, \mathrm{d}z').$$

Notice that $Q_k(g_k R_{k+1:n}) = R_k\, R_{k+1:n} = R_{k:n}$, hence

$$\begin{aligned} \mathrm{var}(g_k R_{k+1:m}1, \mu_{k-1}^{\mathrm{imp}} Q_k) &= \langle \mu_{k-1}^{\mathrm{imp}} Q_k, (g_k R_{k+1:m}1)^2 \rangle - \langle \mu_{k-1}^{\mathrm{imp}} Q_k, (g_k R_{k+1:m}1) \rangle^2 \\ &= \langle \mu_{k-1}^{\mathrm{imp}} Q_k, (g_k R_{k+1:m}1)^2 \rangle - \langle \mu_{k-1}^{\mathrm{imp}}, R_{k:m}1 \rangle^2. \end{aligned}$$

From this equality and using the fact that $g_k^2 = g_k$, one deduces that

$$C_k + D_k = \frac{\langle \gamma_{k-1}^{\mathrm{imp}}, 1 \rangle \langle \gamma_{k-1}^{\mathrm{red}}, 1 \rangle}{\langle \gamma_{k-1}, 1 \rangle^2} \frac{\langle \mu_{k-1}^{\mathrm{imp}} Q_k, g_k\, (R_{k+1:m}1)^2 \rangle}{\langle \mu_{k-1}, R_{k:m}1 \rangle^2} - 1 = \langle \gamma_{k-1}^{\mathrm{red}}, 1 \rangle \frac{\langle \gamma_{k-1}^{\mathrm{imp}} Q_k, g_k\, (R_{k+1:m}1)^2 \rangle}{\langle \gamma_{k-1}, R_{k:m}1 \rangle^2} - 1.$$

Finally, notice that for any bounded measurable function f

$$\begin{aligned}
\langle \gamma_{k-1}^{\mathrm{imp}} Q_k, g_k f \rangle &= \int_{z'} g_k(z') f(z') \int_z \gamma_{k-1}^{\mathrm{imp}}(dz)\, Q_k(z, dz') \\
&= \int_{z'} (g_k(z'))^2 f(z') \int_z \gamma_{k-1}^{\mathrm{imp}}(dz)\, Q_k(z, dz'), \quad \text{since } g_k^2 = g_k \\
&= \int_{z'} g_k^{\mathrm{red}}(z') f(z') \int_z \gamma_{k-1}^{\mathrm{imp}}(dz)\, (g_k^{\mathrm{imp}}(z'))^2 g_k^{\mathrm{red}}(z')\, Q_k(z, dz') \\
&= \int_{z'} g_k^{\mathrm{red}}(z') f(z') \int_z \gamma_{k-1}^{\mathrm{imp}}(dz)\, R_k^{\mathrm{imp}}(z, dz') \\
&= \int_{z'} g_k^{\mathrm{red}}(z') f(z')\, \gamma_k^{\mathrm{imp}}(dz') = \langle \gamma_k^{\mathrm{imp}}, g_k^{\mathrm{red}} f \rangle.
\end{aligned}$$

Thus,

$$C_k + D_k = \langle \gamma_{k-1}^{\mathrm{red}}, 1 \rangle \frac{\langle \gamma_k^{\mathrm{imp}}, g_k^{\mathrm{red}} (R_{k+1:m} 1)^2 \rangle}{\langle \gamma_k, R_{k+1:m} 1 \rangle^2} - 1.$$

□

4 NUMERICAL EXPERIMENTS

The algorithm is tested on the Brownian bridge $\{X_t, 0 \leq t \leq 1\}$, which is solution of the following real stochastic differential equation $\mathrm{d}X_t = X_t/(t-1)\mathrm{d}t + \mathrm{d}W_t$, with $X_0 = 0$. One can show that $X_1 = 0$ almost surely and that

$$\mathbb{P}(T_B \leq 1) = \mathbb{P}(\sup_{0 \leq t \leq 1} X_t \geq B) = \exp^{-2B^2}. \tag{7}$$

Thresholds for the splitting algorithm can easily be chosen adaptively, as explained in section 2.3. According to (Cérou et al. 2006), to choose $p_k = p$ for all $k = 1, \ldots, m$, with p close to 1 is optimal in term of variance reduction. But if p is too close to 1, the number of subsets increases and consequently, the total simulation time increases. According to our experiments, setting $p = 0.6$ is a good trade-off for the Brownian bridge case. In all cases, the number m of intermediate subsets is related to the probability of interest $P = \mathbb{P}(T_B < T)$ with the formula $P = \prod_{k=1}^{m} p_k = p^m$. Thus, $m = \log P / \log p$ and the total number of trajectories of the Markov chain used the algorithm is then in $\mathscr{O}(m \times N)$. The multinomial re-sampling used at each step can be in $\mathscr{O}(N)$. Neglecting the computation of the importance weights and the re-sampling weights, the complexity of the algorithm is then related to the rarity of the event and equals to $\mathscr{O}(m \times N)$. This result stands for I2S algorithm and standard splitting alike. One can also notice that the estimation of thresholds can efficiently be done with a small number of particle. In this article, 100 particles have been used. The $a_p(t)$ functions have to be soft enough to preserve some diversity. According to our experiments, an exponential selection is too strong and does not ensure variance reduction any more. Furthermore, an easy tuning is required to proceed to several tests. For all these reasons, the a_p functions are chosen in the parametric family defined by

$$a_p(t) = a_p^{\alpha}(t) = \frac{1}{t^{\alpha}},$$

for $\alpha \geq 0$. For any real N-sample $(x_1, \ldots, x_N)$, the accuracy of an estimator is measured by its relative standard deviation (*rSTD*), defined as the ratio of the empirical standard deviation to the empirical mean.

Algorithms I2S and standard splitting are unbiased. In figure (1), we compare the rSTD given by the probability estimation of equation (7) with standard splitting algorithm, and I2S algorithm (algorithm 1). We choose $\alpha = 0.05$. I2S is efficient when compared to standard splitting. In figure (2), one presents the evolution of the rSTD in function of parameter α. An optimum value of α minimizing rSTD can be found. Notice that $\alpha = 0$ corresponds to the standard splitting case. The results of figures (1) and (2) have been

obtained with $N = 1000$ particles and over 250 retrials for each α. To understand the gain in probability estimation via I2S algorithm, table 1 compares rSTD of I2S algorithm with parameters $B = 4$, $\alpha = 0.05$ and $N = 5000$ and rSTD of standard splitting for different values of N. Estimations are obtained over 250 retrials.

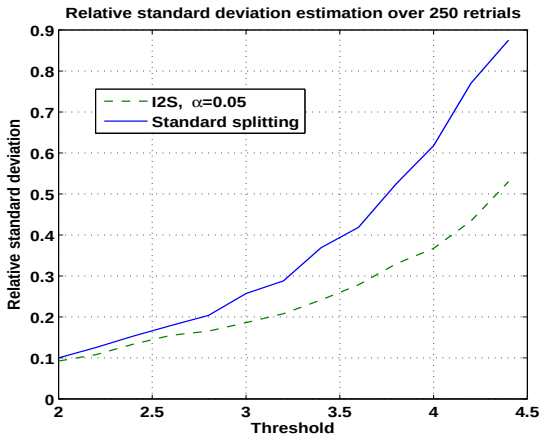

Figure 1: Standard splitting and I2S algorithm.

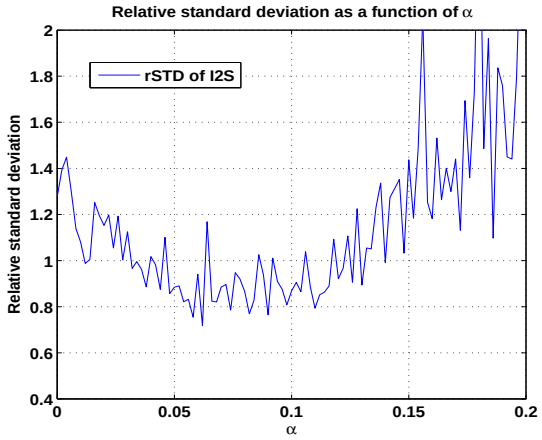

Figure 2: I2S rSTD with different functions $a_p^{\alpha}(t)$.

When one chooses optimal subsets in the standard splitting method, as explained in section 2.3, the theoretical relative standard deviation can be computed (Cérou et al. 2006), and equals to

$$\frac{1}{N}\sum_{k=1}^{m}\left(\frac{1}{p_k}-1\right). \tag{8}$$

We set again $p_k = p = 0.6$ for all $k = 1,\ldots,m$. Table 2 compares the theoretical rSTD of standard splitting with optimal subset *(Opt rSTD)* given by equation (8), the empirical rSTD given by standard splitting with non-optimal subsets, and the rSTD given by I2S algorithm. The number of trajectories sampled at each threshold are the same for both algorithms. One can observe that I2S performances are about halfway the standard splitting with optimal and non-optimal subtsets, in terms of rSTD.

5 COMPARISON WITH EXISTING ALGORITHMS

Standard splitting The importance function Φ can be difficult to choose in an optimal way. A bad choice would lead to inaccurate estimations, and the optimal one depends on the unknown probability to estimate. Our implementation of the proposed algorithm uses a time–independent function Φ already given by the problem, and the intermediate subsets are adaptively chosen, in a non-optimal manner. Precision is increased thanks to the use of the g_k^{imp} and g_k^{red} functions. The total computation time is the same than the one of splitting algorithm without incorporating importance sampling. I2S algorithm can thus improve the accuracy of the algorithm without increasing the total computation time. Notice that I2S algorithm is no longer useful when T is random since the instant when a trajectory enters a new subset does not impact the probability the trajectory has to reach B. Another avenue for research would be to weight and re-sample trajectories with respect to the distance from the rare set.

Del Moral and Garnier algorithm In (Del Moral and Garnier 2005), the authors propose a method to estimate some rare events at a finite horizon time. The probability that they are interested in is $\mathbb{P}(\Phi(Z_m) \in F)$, for Z_k a Markov chain. The following decomposition is used, for any bounded measurable function V :

$$\mathbb{E}(V(Z_m)) = \mathbb{E}(V(Z_m)\prod_{k=1}^{m} G_k^{-}(Z_k)\prod_{k=1}^{m} G_k(Z_k))$$

	I2S	Standard splitting			
N	5000	5000	7500	10000	12500
rSTD	0.36	0.61	0.49	0.47	0.35
time (sec.)	5.55	5.37	8.00	10.66	13.30

Table 1: Convergence of standard splitting.

B	$\mathbb{P}(T_B \leq 1)$	Opt rSTD	Non-optimal splitting		I2S algorithm	
			estimate	time (sec.)	estimate	time (sec.)
2	$3.4.10^{-4}$	0.089	$2.93.10^{-4} \pm 0.10\%$	2.15	$3.00.10^{-4} \pm 0.092\%$	2.12
2.8	$1.5.10^{-7}$	0.11	$1.27.10^{-7} \pm 0.20\%$	3.85	$1.25.10^{-7} \pm 0.17\%$	4.02
3.4	$9.1.10^{-11}$	0.12	$6.72.10^{-11} \pm 0.37\%$	5.60	$6.94.10^{-11} \pm 0.24\%$	5.68
4	$1.3.10^{-14}$	0.13	$8.64.10^{-15} \pm 0.62\%$	7.50	$9.90.10^{-15} \pm 0.36\%$	7.63

Table 2: Comparison of the rSTD given by optimal standard splitting.

where G_p^- and G_p are strictly positive function and $G_p^- G_p = 1$. Setting $V(Z_m) = 1_{\{\Phi(Z_m) \in F\}}$, one obtains the probability of interest. Since the G_p and G_p^- functions are not allowed to be equal to zero, I2S algorithm is an extension the algorithm of Del Moral and Garnier. It is worth noting that, as expected, the expression of the variance in theorem (1) is the same than in (Del Moral and Garnier 2005).

6 CONCLUSION AND PERSPECTIVES

We proposed improvements of the splitting algorithm for rare event based on different selection functions, and on a non-optimal important function. The earlier trajectories reach an intermediate subset, the more often they are selected. Next, a weighted stage in the algorithm is required to prevent probability bias. The a_p functions can not only depend on the time variable, but also on the space variable. Indeed, one can be interested in giving more importance to some space regions, setting $a_p(x,t)$ depending on space variable x or space-time variable (x,t). So the algorithm can be understood as a led exploration of the space state coupling with a rare event estimation. Then, we established a central limit theorem for the estimation of the rare event probability. For well-known toy cases, a numerical or analytical computation of the variance given in theorem 1 is possible, provided explicit density of the random vector $(X_{T_1}, T_1, \ldots, X_{T_m}, T_m)$ is established. It could give a better idea on how to choose the optimal functions a_p. Numerical experiments showed variance reductions compared with the standard splitting method. Beyond the pedagogical side of this paper, a mid-term objective could be to apply this algorithm to the estimation of conflict probability between aircraft, using models such those developed in (Prandini, Blom, and Bakker 2011).

ACKNOWLEDGEMENT

The work of Damien Jacquemart-Tomi is financially supported by DGA (Direction Générale de l'Armement) and Onera, The French Aerospace Lab.

REFERENCES

Cérou, F., P. Del Moral, A. Guyader, F. Le Gland, P. Lezaud, and H. Topart. 2006, September. "Some recent improvements to importance splitting". In *6th International Workshop on Rare Event Simulation*. Bamberg, Germany.

Cérou, F., P. Del Moral, F. Le Gland, and P. Lezaud. 2005. "Limit theorems for the multilevel splitting algorithm in the simulation of rare events". In *Proceedings of the 2005 Winter Simulation Conference*, edited by M. E. Kuhl, N. M. Steiger, F. B. Armstrong, and J. A. Joines, 682–691. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Cérou, F., P. Del Moral, F. Le Gland, and P. Lezaud. 2006. "Genetic Genealogical Models in Rare Event Analysis". *ALEA* 1:1–19.

Cérou, F., and A. Guyader. 2007. "Adaptive Multilevel Splitting for Rare Event Analysis". *Stochastic Analysis and Applications* 25 (2): 417–443.

Del Moral, P., and J. Garnier. 2005. "Genealogical particle analysis of rare events". *The Annals of Applied Probability* 15 (4): 2496–2534.

Garvels, M. 2000. *The splitting in rare event estimation*. Ph.D. thesis, University of Twente, Twente, Netherlands. Available via http://doc.utwente.nl/29637/1/t0000013.pdf [accessed July 25, 2013].

Garvels, M., J. Van Ommeren, and D. Kroese. 2002. "On the importance function in splitting simulation". *European Transactions on Telecommunications* 13 (4): 363–371.

Glasserman, P., P. Heidelberger, P. Shahabuddin, and T. Zajic. 1998. "A Large Deviations perspective on the Efficiency of Multilevel Splitting". *IEEE Transactions on Automatic Control* AC-43 (12): 1666–1679.

Jacquemart, D., and J. Morio. 2013, January. "Conflict Probability Estimation Between Aircraft with Dynamic Importance Splitting". *Safety Science* 51 (1): 94–100.

Juneja, S., and P. Shahabuddin. 2006. "Rare-Event Simulation Techniques : An Introduction and Recent Advances". *Handbooks in Operations Research and Management Science* 13 (06): 291–350.

Le Gland, F. 2007. "Combined use of importance weights and resampling weights in sequential Monte Carlo methods". *ESAIM: Proceedings* 19:85–100.

L'Ecuyer, P., V. Demers, and B. Tuffin. 2006. "Splitting for Rare-Event Simulation". In *Proceedings of the 2006 Winter Simulation Conference, Monterey*, edited by L. F. Perrone, B. G. Lawson, J. Liu, and F. P. Wieland, 137–148. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

L'Écuyer, P., V. Demers, and B. Tuffin. 2007, April. "Rare events, splitting, and quasi–Monte Carlo". *ACM Transactions on Modeling and Computer Simulation* 17 (2, Special issue honoring Perwez Shahabuddin). Article 9.

L'Écuyer, P., F. Le Gland, P. Lezaud, and B. Tuffin. 2009. "Splitting methods". In *Monte Carlo Methods for Rare Event Analysis*, edited by G. Rubino and B. Tuffin, Chapter 3, 39–61. Chichester: John Wiley & Sons.

Prandini, M., H. Blom, and G. Bakker. 2011. "Air traffic complexity and the interacting particle system method: An integrated approach for collision risk estimation". In *American Control Conference (ACC)*, 2154–2159. San Francisco, California.

AUTHOR BIOGRAPHIES

DAMIEN JACQUEMART-TOMI is a Ph.D student in applied mathematics at ONERA and INRIA. He obtained agrégation in mathematics, in 2010. Then, he graduated from the University of Marseille, France with a Research Masters Degree in Probability and Statistics, in 2011. His research interests include rare event estimation with aerospace applications. His email address is damien.jacquemart@onera.fr.

FRANCOIS LE GLAND is a DR (directeur de recherche) at INRIA Rennes since 1993, where he is the head of the ASPI research team. He graduated from Ecole Centrale Paris in 1978 and obtained a Ph.D. in applied mathematics from université Paris Dauphine in 1981. His main research theme is the application of interacting particle systems to Bayesian estimation, to rare event estimation and to global optimization. His email address is francois.le_gland@inria.fr.

JÉRÔME MORIO is a Research Engineer at ONERA since 2007. He graduated from Ecole Nationale Supérieure de Physique de Marseille, France, in 2004 and obtained a Ph.D. in physics from Aix-Marseille University, France, in 2007. His main research interests include rare event estimations, sensitivity analysis and uncertainty propagation. His email address is jerome.morio@onera.fr.