OPTIMAL INPUT POTENTIAL FUNCTIONS IN THE INTERACTING PARTICLE SYSTEM METHOD

H Chraibi, A Dutfoy, T Galtier, J. Garnier

Applied Probability Trust (14 November 2018)

OPTIMAL INPUT POTENTIAL FUNCTIONS IN THE INTERACTING PARTICLE SYSTEM METHOD

H. CHRAIBI,* *Électricité de France (EDF), PERICLES department*

A. DUTFOY,** *Électricité de France (EDF), PERICLES department*

T. GALTIER,*** *Universit Paris-Diderot, LPSM*

J. GARNIER,**** *École polytechnique, CMAP*

Abstract

The assessment of the probability of a rare event with a naive Monte-Carlo method is computationally intensive, so faster estimation methods, such as variance reduction methods, are needed. We focus on one of these methods which is the interacting particle (IPS) system method.

The method requires to specify a set of potential functions. The choice of these functions is crucial, because it determines the magnitude of the variance reduction. So far, little information was available on how to choose the potential functions. To remedy this, we provide the expression of the optimal potential functions minimizing the asymptotic variance of the estimator of the IPS method.

Keywords: Rare event ; Reliability assessment ; Feymann-Kac particle filters; interacting particles system ; Sequential Monte-Carlo Samplers ; selection ; potential function

2010 Mathematics Subject Classification: Primary 60K10

Secondary 90B25;62N05

* Postal address: EDF lab saclay, 7 Bd Gaspard Monge, 91120 Palaiseau, France

** Postal address: EDF lab saclay, 7 Bd Gaspard Monge, 91120 Palaiseau, France

*** Postal address: Laboratoire de probabilités, statistique et modélisation (LPSM), 75205 Paris Cedex 13, France

** Email address: tgaltier@gmail.com

**** Postal address: Centre de mathématiques appliquées (CMAP), 91128 Palaiseau Cedex, France

1. Introduction

Let $(Z_k)_{k\in\{0,\dots,n\}}$ be a Markov chain with values in the measurable spaces $(E_k, \mathcal{E}_k)$, and with initial law ν_0, and with a kernel Q_k such that for $k > 0$ and for any bounded measurable function $h : E_k \to R$

$$\mathbb{E}[h(Z_k)|Z_{k-1}] = \int_{E_k} h(z_k)Q_k(dz_k|Z_{k-1}).$$

For any bounded measurable function $h : E_0 \times \cdots \times E_n \to R$ we have:

$$\mathbb{E}[h(Z_0,\dots,Z_n)] = \int_{E_n\times\cdots\times E_0} h(z_0,\dots,z_n)Q_n(dz_n|z_{n-1})\cdots Q_1(dz_1|z_0)\nu_0(dz_0)$$

Let $\mathbf{Z}_k = (Z_0, Z_1, \dots, Z_k)$ be a trajectory of size k, and let $\mathbf{E}_k = E_0 \times E_1 \times \cdots \times E_k$ be the set of trajectories of size k that we equip with the product σ-algebra $\boldsymbol{\mathcal{E}}_k = \mathcal{E}_0 \otimes \mathcal{E}_1 \otimes \cdots \otimes \mathcal{E}_k$. For $i < j$, consider two trajectories $\mathbf{z}_i$ and $\mathbf{z}_j$: when it is necessary to differentiate the coordinates of these trajectories we write the coordinates $z_{i,k}$ for $k \le i$ and $z_{j,k}$ for $k \le j$ such that $\mathbf{z}_i = (z_{i,0}, z_{i,1}, \dots, z_{i,i})$ and $\mathbf{z}_j = (z_{j,0}, z_{j,1}, \dots, z_{j,j})$. We introduce the Markov Chain of the trajectories $(\mathbf{Z}_k)_{k\ge0}$ with values in the measurable spaces $(\mathbf{E}_k, \boldsymbol{\mathcal{E}}_k)$, and with the transition kernels M_k such that:

$$M_k(d\mathbf{z}_k|\mathbf{z}_{k-1}) = \delta_{\mathbf{z}_{k-1}}\big(d(z_{k,1},\ \dots\ , z_{k,k-1})\big)Q_k(dz_k|z_{k-1,k-1}).$$

For any bounded measurable function $h : \mathbf{E}_n \to \mathbb{R}$ we have

$$p_h = \mathbb{E}\left[h(\mathbf{Z}_n)\right] = \iiint_{\mathbf{E}_n\times\cdots\times\mathbf{E}_0} h(\mathbf{z}_n)\prod_{k=1}^{n} M_k(d\mathbf{z}_k|\mathbf{z}_{k-1})\nu_0(d\mathbf{z}_0). \tag{1}$$

The IPS method provides an estimator $\hat{p}_h$ of p_h with a different variance than the Monte-Carlo estimator. It was first introduced in [10], and with an alternative formulation in [9]. The IPS takes in input potential functions, denoted $G_k : \mathbf{E}_k \to \mathbb{R}^+$ for $k \le n$. When these potential functions are properly tuned, the method can yield an estimator with a smaller variance than the Monte-Carlo estimator. When such is the case, the IPS method can assess the quantity p_h with significantly less simulation runs than the Monte-Carlo method, for the same accuracy. Such variance reduction method is recommended in rare event analysis, where the Monte-Carlo method is well-known to be computationally intensive.

Rare event analyses are often carried out for reliability assessment where we want to assess the probability of failure of a reliable system. Typically there is a region D of $\mathbf{E}_n$ that corresponds to the system failure, and we want to assess the probability of failure which happens when $\mathbf{Z}_n$ enters D. In order to assess this probability of failure $\mathbb{P}(\mathbf{Z}_n \in D)$, one can take $h(\mathbf{Z}_n) = \mathbb{1}_D(\mathbf{Z}_n)$ so that $p_h = \mathbb{P}(\mathbf{Z}_n \in D)$ and use the IPS method to get the estimation $\hat{p}_h$. Although the main application of the IPS method is reliability assessment where it relates to the case $h = \mathbb{1}_D$, the result of this paper will be presented in its most general form, where h is an arbitrary measurable function.

As we said, the choice of the potential functions $(G_k)_{k<n}$ is paramount because it determines the variance of the IPS estimator, but so far, little information has been provided on the form of efficient potential functions. The standard approach is to find the best potential function within a set of parametric potential functions, and so the efficiency of the method strongly depends on the quality of the chosen parametric family. For instance, in [10] the authors obtain their best variance reduction by choosing

$$G_k(\mathbf{Z}_k) = \frac{\exp\left[-\lambda V(Z_k)\right]}{\exp\left[-\lambda V(Z_{k-1})\right]}$$

where λ is a positive tuning parameter, and the quantity $V(z) = a - z$ roughly measures the proximity of z to the critical region that was $D = [a; +\infty)$. In [15] they stress out that it seems better to take a time-dependent proximity function V_k instead of V, yielding:

$$G_k(\mathbf{Z}_k) = \frac{\exp\left[-\lambda V_k(Z_k)\right]}{\exp\left[-\lambda V_{k-1}(Z_{k-1})\right]},$$

where the quantities $V_k(z)$ are again measuring the proximity of z to D. Once the set of parametric potential functions is chosen, it is necessary to optimize the tuning parameters of the potentials. Different methods have been proposed. In [13], an empirical heuristic algorithm is provided; in [12] a meta model of the variance is minimized; in [10] the large deviation principle is used as a guide. One other common option for the potential functions is the one done in splitting methods. Indeed the splitting method can also be seen as a version of the IPS method [5]. In this method one wants to assess the probability that a random variable Z belongs to a subset B_n. A succession of nested sets $E = B_0 \supseteq B_1 \supseteq B_2 \supseteq \cdots \supseteq B_n$ is chosen by the practitioner or possibly chosen in an adaptive manner [4]. One considers a sequence ofrandom

variables $(Z_i)_{i=1..n}$ such that the small probability $\mathbb{P}(Z \in B_n)$ can be decomposed into a product of conditional probabilities: $\mathbb{P}(Z \in B_n) = \prod_{i=1}^{n} \mathbb{P}(Z_i \in B_i | Z_{i-1} \in B_{i-1})$ and $\mathbb{P}(Z \in B_n) = \mathbb{E}\left[h(\mathbf{Z}_n)\right]$ by setting $h(\mathbf{Z}_n) = \mathbb{1}_{B_n}(Z_n)$. In this method the potential functions are chosen of the following form

$$G_k(\mathbf{Z}_k) = \mathbb{1}_{B_k}(Z_k).$$

One usually optimizes the variance reduction within this family of potential functions by optimizing the choice of the sets $(B_k)_{k \leq n}$.

In this paper we tackle the issue of the choice of the potential functions. Our contribution is to provide the expressions of the theoretical optimal potential functions that minimize the variance of the estimator of the IPS method. We hope these expressions will lead the practitioners to design more efficient potential functions, that are closer from the optimal ones.

The rest of the paper is organized as follows. The section 2 introduces the IPS method, and section 3 presents the potential functions that minimize the asymptotic variance of the IPS estimator, then the section 4 presents two example of applications on a toy model, and finally section 5 discusses the implications of our results.

In the rest of the paper we use the following notations: We denote by $\mathcal{M}(A)$ the set of bounded measurable functions on a set A. If f is a bounded measurable function, and η is a measure we note $\eta(f) = \int f \, d\eta$. If M is a Markovian kernel, we denote by $M(f)$ the function such that $M(f)(x) = \int f(y) M(dy|x)$, and for a measure η, we denote by ηM the measure such that

$$\eta M(f) = \int \int f(y) M(dy|x) \eta(dx).$$

2. The IPS method

2.1. A Feynman-Kac model

The IPS method relies on a Feynman-Kac model [8] which is defined in this subsection. For each $k \leq n$, we define a *target* probability measure $\tilde{\eta}_k$ on $(\mathbf{E}_k, \boldsymbol{\mathcal{E}}_k)$, such that:

$$\forall B \in \boldsymbol{\mathcal{E}}_k, \quad \tilde{\eta}_k(B) = \frac{\mathbb{E}\left[\mathbb{1}_B(\mathbf{Z}_k)\prod_{s=0}^{k} G_s(\mathbf{Z}_s)\right]}{\mathbb{E}\left[\prod_{s=0}^{k} G_s(\mathbf{Z}_s)\right]}. \tag{2}$$

For each $k,\ 1 \leq k \leq n$, we define the *propagated* target probability measure η_k on $(\mathbf{E}_k, \boldsymbol{\mathcal{E}}_k)$ such that $\eta_k = \tilde{\eta}_{k-1}M_{k-1}$ and $\eta_0 = \tilde{\eta}_0$. We have:

$$\forall B \in \boldsymbol{\mathcal{E}}_{k+1}, \quad \eta_{k+1}(B) = \frac{\mathbb{E}\left[\mathbb{1}_B(\mathbf{Z}_{k+1})\prod_{s=0}^{k} G_s(\mathbf{Z}_s)\right]}{\mathbb{E}\left[\prod_{s=0}^{k} G_s(\mathbf{Z}_s)\right]}. \tag{3}$$

Let Ψ_k be the application that transforms a measure η defined on $\mathbf{E}_k$ into a measure $\Psi_k(\eta)$ defined on $\mathbf{E}_k$ and such that

$$\Psi_k(\eta)(f) = \frac{\int G_k(\mathbf{z})f(\mathbf{z})\eta(d\mathbf{z})}{\eta(G_k)}. \tag{4}$$

We say $\Psi_k(\eta)$ gives the selection of η through the potential G_k. Notice that $\tilde{\eta}_k$ is the selection of η_k as $\tilde{\eta}_k = \Psi_k(\eta_k)$. The target distributions can therefore be built according to the following pattern of successive selection and propagation steps:

$$\eta_k \xrightarrow{\Psi_k} \tilde{\eta}_k \xrightarrow{.M_k} \eta_{k+1}.$$

We also define the associated unnormalized measures $\tilde{\gamma}_k$ and γ_{k+1}, such that for $f \in \mathcal{M}(\mathbf{E}_k)$:

$$\tilde{\gamma}_k(f) = \mathbb{E}\left[f(\mathbf{Z}_k)\prod_{s=0}^{k} G_s(\mathbf{Z}_s)\right] \quad \text{and} \quad \tilde{\eta}_k(f) = \frac{\tilde{\gamma}_k(f)}{\tilde{\gamma}_k(1)}, \tag{5}$$

and for $f \in \mathcal{M}(\mathbf{E}_{k+1})$:

$$\gamma_{k+1}(f) = \mathbb{E}\left[f(\mathbf{Z}_{k+1})\prod_{s=0}^{k} G_s(\mathbf{Z}_s)\right] \quad \text{and} \quad \eta_{k+1}(f) = \frac{\gamma_{k+1}(f)}{\gamma_{k+1}(1)}. \tag{6}$$

Denoting $f_h(\mathbf{z}) = \frac{h(\mathbf{z})}{\prod_{s=0}^{n-1} G_s(\mathbf{z})}$, notice that we have:

$$p_h = \gamma_n(f_h) = \eta_n(f_h)\prod_{k=0}^{n-1}\eta_k\big(G_k\big). \tag{7}$$

2.2. The IPS's algorithm and its estimator

The IPS method provides an algorithm to generate samples whose weighted empirical measures approximate the probability measures η_k and $\tilde{\eta}_k$ respectively for each step k. These approximations are then used to provide an estimator of p_h. For the sample approximating η_k, we denote $\mathbf{Z}_k^j$ the j^{th} trajectory and W_k^j its weight. Respectively in the sample approximating $\tilde{\eta}_k$, we denote $\widetilde{\mathbf{Z}}_k^j$ the j^{th} trajectory and $\widetilde{W}_k^j$ its associated weight. For simplicity reasons, in this paper, we consider that the samples all contain N trajectories, but it is possible to modify the sample size at each step, as illustrated in [14]. The empirical measure approximating η_k and $\tilde{\eta}_k$ are denoted by η_k^N and $\tilde{\eta}_k^N$ and are defined by:

$$\tilde{\eta}_k^N = \sum_{i=1}^{N} \widetilde{W}_k^i \, \delta_{\widetilde{\mathbf{Z}}_k^i} \quad \text{and} \quad \eta_k^N = \sum_{i=1}^{N} W_k^i \, \delta_{\mathbf{Z}_k^i} \, . \tag{8}$$

So for all $k \leq n$ and $f \in \mathcal{M}(\mathbf{E}_k)$,

$$\tilde{\eta}_k^N(f) = \sum_{i=1}^{N} \widetilde{W}_k^i \, f\big(\widetilde{\mathbf{Z}}_k^i\big) \quad \text{and} \quad \eta_k^N(f) = \sum_{i=1}^{N} W_k^i \, f\big(\mathbf{Z}_k^i\big) \, . \tag{9}$$

By plugging these estimations into equations (5) and (6), we get estimations for the unnormalized distributions. Denoting by $\tilde{\gamma}_k^N$ and γ_k^N these estimations, for all $k \leq n$ and $f \in \mathcal{M}(\mathbf{E}_k)$, we have:

$$\tilde{\gamma}_k^N(f) = \tilde{\eta}_k^N(f) \prod_{s=0}^{k-1} \eta_s^N(G_s) \quad \text{and} \quad \gamma_k^N(f) = \eta_k^N(f) \prod_{s=0}^{k-1} \eta_s^N(G_s). \tag{10}$$

In particular if we apply (10) to the test function f_h, we get an estimator $\hat{p}_h$ of p_h defined by:

$$\hat{p}_h = \eta_n^N(f_h) \prod_{k=0}^{n-1} \eta_k^N\big(G_k\big). \tag{11}$$

The IPS algorithm builds the samples sequentially, alternating between a selection step and a propagation step.

The k^{th} selection step transforms the sample $(\mathbf{Z}_k^j, W_k^j)_{j \leq N}$, into the sample $(\widetilde{\mathbf{Z}}_k^j, \widetilde{W}_k^j)_{j \leq N}$. This transformation is done with a multinomial resampling scheme. This means that the $\widetilde{\mathbf{Z}}_k^j$'s are drawn with replacement from the sample $(\mathbf{Z}_k^j)_{j \leq N}$, each

Initialization: $k=0,\ \forall j=1..N,\ \mathbf{Z}_0^j \overset{i.i.d.}{\sim} \eta_0$ and $W_0^j=\frac{1}{N}$, and $\tilde{W}_0^j=\frac{G_0(\mathbf{Z}_0^j)}{\sum_s G_0(\mathbf{Z}_0^s)}$

while $k<n$ **do**

Selection:

$(\tilde{N}_k^j)_{j=1..N} \sim Mult\big(N,(\tilde{W}_k^j)_{j=1..N}\big)$

$\forall j:=1..N,\ \tilde{W}_k^j:=\frac{1}{N}$

Propagation:

for $j:=1..N$ **do**

using the kernel M_k, continue the trajectory $\tilde{\mathbf{Z}}_k^j$ to get $\mathbf{Z}_{k+1}^j$

set $W_{k+1}^j=\tilde{W}_k^j$ and $\tilde{W}_{k+1}^j=\frac{W_{k+1}^j G_{k+1}(\mathbf{Z}_{k+1}^j)}{\sum_s W_{k+1}^s G_{k+1}(\mathbf{Z}_{k+1}^s)}$

if $\forall j,\ \tilde{W}_{k+1}^j=0$ **then**

$\forall q>k$, set $\eta_q^N=\tilde{\eta}_q^N=0$ and Stop

else

$k:=k+1$

TABLE 1: IPS algorithm

trajectory $\mathbf{Z}_k^j$ having a probability $\frac{W_k^j G_k(\mathbf{Z}_k^j)}{\sum_{i=1}^N W_k^i G_k(\mathbf{Z}_k^i)}$ to be drawn each time. We let $\tilde{N}_k^j$ be the number of times the particle $\mathbf{Z}_k^j$ is replicated in the sample $(\tilde{\mathbf{Z}}_k^j,\tilde{W}_k^j)_j$, so $N=\sum_{j=1}^N \tilde{N}_k^j$. After this resampling the weights $\tilde{W}_k^j$ are set to $\frac{1}{N}$.

The interest of this selection by resampling is that it discards low potential trajectories and replicates high potential trajectories. Thus, the selected sample focuses on trajectories that will have a greater impact on the estimations of the next distributions once extended.

If one specifies potential functions that are not positive everywhere, there can be a possibility that at a step k we get $\forall j, G_k(\mathbf{Z}_k^j)=0$. When this is the case, the probability for resampling can not be defined, the algorithm stops, and we consider that $\forall s\geq k$ the measures $\tilde{\eta}_s^N$ and η_{s+1}^N are equal to the null measure.

Then the k^{th} propagation step transforms the sample $(\tilde{\mathbf{Z}}_k^j,\tilde{W}_k^j)_{j\leq N}$, into the sample $(\mathbf{Z}_{k+1}^j,W_{k+1}^j)_{j\leq N}$. Each trajectory $\mathbf{Z}_{k+1}^j$ is obtained by extending the trajectory $\tilde{\mathbf{Z}}_k^j$ on step further using the transition kernel M_k. The weights satisfy $W_{k+1}^j=\tilde{W}_k^j,\ \forall j$. Then the procedure is iterated until the step n. The full algorithm to build the samples is displayed in table 1.

For $k < n$, we denote by $\hat{\mathbf{E}}_k = \{\mathbf{z}_k \in \mathbf{E}_k, G_k(\mathbf{z}_k) > 0\}$ the support of G_k, and we denote $\hat{\mathbf{E}}_n = \{\mathbf{z}_n \in \mathbf{E}_n, h(\mathbf{z}_n) > 0\}$ the support of h. We will make the following assumption on the potential functions:

$$\exists\, \varepsilon > 0, \quad \forall k \leq n,\ \forall \mathbf{z}_{k-1} \in \hat{\mathbf{E}}_{k-1},\ M_{k-1}(\hat{\mathbf{E}}_k | \mathbf{z}_{k-1}) > \varepsilon, \tag{G}$$

Theorem 1. *When the potential functions satisfy* (G)*, $\hat{p}_h$ is unbiased and strongly consistent.*

The proof of theorem 1 can be found in [8] chapter 7.

Theorem 2. *When the potential are strictly positive:*

$$\sqrt{N}(\hat{p}_h - p_h) \xrightarrow[N\to\infty]{d} \mathcal{N}\big(0, \sigma^2_{IPS,G}\big) \tag{12}$$

where, with the convention that $\prod_{i=0}^{-1} G_i(\mathbf{Z}_i) = \prod_{i=0}^{-1} G_i^{-1}(\mathbf{Z}_i) = 1$*:*

$$\sigma^2_{IPS,G} = \sum_{k=0}^{n} \left\{ \mathbb{E}\Big[\prod_{i=0}^{k-1} G_i(\mathbf{Z}_i) \Big] \mathbb{E}\Big[\mathbb{E}[h(\mathbf{Z}_n)|\mathbf{Z}_k]^2 \prod_{s=0}^{k-1} G_s^{-1}(\boldsymbol{Z}_s) \Big] - p_h^2 \right\}. \tag{13}$$

A proof of this CLT can be found in [8] chapter 9. This CLT is an important result of the particle filter literature. The non-asymptotic fluctuations of the particle filters as been studied in [3]. Recently two weakly consistent estimators of the asymptotic variance $\sigma^2_{IPS,G}$, that are based on a single run of the method, have been proposed in [14]. One of these estimators is closely related to what was proposed in [6].

3. The theoretical expression for the optimal potential

Here we aim at estimating p_h, the finality is not the estimation of some distributions η_k and $\widetilde{\eta}_k$ for $k < n$, so we can choose any potential functions $(G_k)_{k<n}$ providing they are positive. But the choice of the potential functions has an impact on the variance of the estimation, so we would like to find potential functions that minimize the asymptotic variance (13). Also, note that if potential functions G_k and G'_k are such that $G_k = a.G'_k$ with $a > 0$, then they yield the same variance: $\sigma^2_{IPS,G} = \sigma^2_{IPS,G'}$. Therefore all potential functions will be defined up to a multiplicative term.

Theorem 3. *For $k \geq 1$, let G_k^* be defined by:*

$$G_k^*(\mathbf{Z}_k) \propto \begin{cases} \sqrt{\dfrac{\mathbb{E}\Big[\ \mathbb{E}\big[h(\mathbf{Z}_n)\big|\mathbf{Z}_{k+1}\big]^2\Big|\mathbf{Z}_k\Big]}{\mathbb{E}\Big[\ \mathbb{E}\big[h(\mathbf{Z}_n)\big|\mathbf{Z}_k\ \big]^2\Big|\mathbf{Z}_{k-1}\Big]}} & \text{if } \mathbb{E}\Big[\mathbb{E}\big[h(\mathbf{Z}_n)\big|\mathbf{Z}_k\big]^2\big|\mathbf{Z}_{k-1}\Big] \neq 0 \\ 0 & \text{if } \mathbb{E}\Big[\mathbb{E}\big[h(\mathbf{Z}_n)\big|\mathbf{Z}_k\big]^2\big|\mathbf{Z}_{k-1}\Big] = 0 \end{cases} \tag{14}$$

and for $k = 0$,

$$G_0^*(\mathbf{Z}_0) \propto \sqrt{\mathbb{E}\Big[\mathbb{E}\big[h(\mathbf{Z}_n)\big|\mathbf{Z}_1\big]^2\big|\mathbf{Z}_0\Big]}. \tag{15}$$

The potential functions minimizing $\sigma^2_{IPS,G}$ are the ones that are proportional to the G_k^*'s $\forall k \leq n$. The optimal variance of the IPS method with n steps is then

$$\begin{aligned} \sigma^2_{IPS,G^*} &= \mathbb{E}\left[\mathbb{E}\big[h(\mathbf{Z}_n)\big|\mathbf{Z}_0\big]^2\right] - p_h^2 \\ &\quad + \sum_{k=1}^{n}\left\{\mathbb{E}\left[\sqrt{\mathbb{E}\Big[\ \mathbb{E}\big[h(\mathbf{Z}_n)\big|\mathbf{Z}_k\big]^2\Big|\mathbf{Z}_{k-1}\Big]}\right]^2 - p_h^2\right\}. \end{aligned} \tag{16}$$

Proof. As we lack mathematical tools to minimize $\sigma^2_{IPS,G}$ over the set of positive functions $(G_k)_{k\leq n}$, we had to guess the expressions (14) and (15) before providing the proof of the results. We begin this proof by presenting the heuristic reasoning that provided the expressions (14) and (15).

Assuming we already know the $k-2$ first potential functions, we started by trying to find the $k-1$-th potential function G_{k-1} that minimizes the k-th term of the sum in (13). This is equivalent to minimize the quantity

$$\mathbb{E}\Big[\prod_{i=0}^{k-1} G_i(\mathbf{Z}_i)\Big]\mathbb{E}\Big[\mathbb{E}\big[\mathbb{E}[h(\mathbf{Z}_n)|\mathbf{Z}_k]^2\big|\mathbf{Z}_{k-1}\big]\prod_{s=0}^{k-1} G_s^{-1}(\mathbf{Z}_s)\Big] \tag{17}$$

over G_{k-1}. As the G_{k-1} are equivalent up to a multiplicative constant, we simplify the equation by choosing a multiplicative constant so that $\mathbb{E}\Big[\prod_{i=0}^{k-1} G_i(\mathbf{Z}_i)\Big] = 1$. Our minimizing problem then becomes the minimization of (17) under the constraint $\mathbb{E}\Big[\prod_{i=0}^{k-1} G_i(\mathbf{Z}_i)\Big] = 1$. In order to be able to use a Lagrangian minimization we temporarily assume that the distribution of $\mathbf{Z}_{k-1}$ is discrete and that $\mathbf{Z}_{k-1}$ takes its values in a finite or numerble set $\mathbf{E}$. For $\mathbf{z} \in \mathbf{E}$, we denote $a_{\mathbf{z}} = \mathbb{P}(\mathbf{Z}_{k-1} = \mathbf{z})$ and $d_{\mathbf{z}} = \mathbb{E}\big[\mathbb{E}[h(\mathbf{Z}_n)|\mathbf{Z}_k]^2\big|\mathbf{Z}_{k-1} = \mathbf{z}\big]$ and $g_{\mathbf{z}} = \prod_{i=0}^{k-2} G_i(\mathbf{Z}_i)G_{k-1}(\mathbf{z})$ our minimization problem becomes the minimization of

$$\mathcal{L} = \left(\sum_{\mathbf{z}\in\mathbf{E}} \frac{p_{\mathbf{z}} d_{\mathbf{z}}}{g_{\mathbf{z}}}\right) - \lambda\left(1 - \sum_{\mathbf{z}\in\mathbf{E}} p_{\mathbf{z}} g_{\mathbf{z}}\right) \tag{18}$$

Finding the minimum of this Lagrangian we get that $g_{\mathbf{z}} = \frac{\sqrt{d_{\mathbf{z}}}}{\sum_{\mathbf{z}' \in \mathbf{E}} p_{\mathbf{z}'} \sqrt{d_{\mathbf{z}'}}}$. Now relaxing the constraint of the multiplicative constant, we get that

$$\prod_{i=0}^{k-1} G_i(\mathbf{Z}_i) \propto \sqrt{\mathbb{E}\big[\mathbb{E}[h(\mathbf{Z}_n)|\mathbf{Z}_k]^2 \big| \mathbf{Z}_{k-1}\big]},$$

which gives the desired expressions. After these heuristic arguments we can now rigorously check that these expressions, obtained by minimizing each of the term of sum in (13) one by one, also minimize the whole sum for any distribution of the $\mathbf{Z}_{k-1}$'s.

The proof now consists in showing that, for any set of potential functions (G_s), we have $\sigma^2_{IPS,G} \geq \sigma^2_{IPS,G^*}$. This is done by bounding from below each term of the sum in (13).

We start by decomposing a product of potential functions as follows:

$$\forall k \in \{1, \dots, n\}, \qquad \prod_{s=0}^{k-1} G_s(\mathbf{Z}_s) = \epsilon_{k-1}(\mathbf{Z}_{k-1}) \prod_{s=0}^{k-1} G^*_s(\mathbf{Z}_s) + \bar{\epsilon}_{k-1}(\mathbf{Z}_{k-1}) \tag{19}$$

where when $\mathbf{Z}_{k-1} \in supp \prod_{s=0}^{k-1} G^*_s$,

$$\epsilon_{k-1}(\mathbf{Z}_{k-1}) = \frac{\prod_{s=0}^{k-1} G_s(\mathbf{Z}_s)}{\prod_{s=0}^{k-1} G^*_s(\mathbf{Z}_s)}, \text{ and } \bar{\epsilon}_{k-1}(\mathbf{Z}_{k-1}) = 0$$

and when $\mathbf{Z}_{k-1} \notin supp \prod_{s=0}^{k-1} G^*_s$,

$$\epsilon_{k-1}(\mathbf{Z}_{k-1}) = 0, \text{ and } \bar{\epsilon}_{k-1}(\mathbf{Z}_{k-1}) = \prod_{s=0}^{k-1} G_s(\mathbf{Z}_s).$$

Using (19) we get that

$$\begin{aligned}
&\mathbb{E}\Big[\prod_{s=0}^{k-1} G_s(\mathbf{Z}_s)\Big] \mathbb{E}\Big[\mathbb{E}[h(\mathbf{Z}_n)|\mathbf{Z}_k]^2 \prod_{s=0}^{k-1} G_s^{-1}(\mathbf{Z}_s)\Big] \\
&= \mathbb{E}\Big[\epsilon_{k-1}(\mathbf{Z}_{k-1}) \prod_{s=0}^{k-1} G^*_s(\mathbf{Z}_s)\Big] \mathbb{E}\Big[\mathbb{E}\big[\mathbb{E}[h(\mathbf{Z}_n)|\mathbf{Z}_k]^2 \big| \mathbf{Z}_{k-1}\big] \prod_{s=0}^{k-1} G_s^{-1}(\mathbf{Z}_s)\Big] \\
&\quad + \mathbb{E}\Big[\bar{\epsilon}_{k-1}(\mathbf{Z}_{k-1})\Big] \mathbb{E}\Big[\mathbb{E}\big[\mathbb{E}[h(\mathbf{Z}_n)|\mathbf{Z}_k]^2 \big| \mathbf{Z}_{k-1}\big] \prod_{s=0}^{k-1} G_s^{-1}(\mathbf{Z}_s)\Big] \\
&\geq \mathbb{E}\Big[\epsilon_{k-1}(\mathbf{Z}_{k-1}) \prod_{s=0}^{k-1} G^*_s(\mathbf{Z}_s)\Big] \mathbb{E}\left[\frac{\mathbb{E}\big[\mathbb{E}[h(\mathbf{Z}_n)|\mathbf{Z}_k]^2 \big| \mathbf{Z}_{k-1}\big]}{\prod_{s=0}^{k-1} G_s(\mathbf{Z}_s)}\right] + 0
\end{aligned} \tag{20}$$

For $\mathbf{Z}_{k-1} \in supp \prod_{s=0}^{k-1} G_s^*$ we have:

$$\prod_{s=0}^{k-1} G_s^*(\mathbf{Z}_s) \propto \sqrt{\mathbb{E}\Big[\mathbb{E}\big[h(\mathbf{Z}_n)\big|\mathbf{Z}_k\big]^2 \big|\mathbf{Z}_{k-1}\Big]}$$

So $supp\, \mathbb{E}\big[\mathbb{E}[h(\mathbf{Z}_n)|\mathbf{Z}_k]^2\big|\mathbf{Z}_{k-1}\big] = supp \prod_{s=0}^{k-1} G_s^*$ and we get

$$\begin{aligned}\mathbb{E}\left[\frac{\mathbb{E}\big[\mathbb{E}[h(\mathbf{Z}_n)|\mathbf{Z}_k]^2\big|\mathbf{Z}_{k-1}\big]}{\prod_{s=0}^{k-1} G_s(\mathbf{Z}_s)}\right] &= \mathbb{E}\left[\frac{\mathbb{E}\big[\mathbb{E}[h(\mathbf{Z}_n)|\mathbf{Z}_k]^2\big|\mathbf{Z}_{k-1}\big]}{\epsilon_{k-1}(\mathbf{Z}_{k-1}) \prod_{s=0}^{k-1} G_s^*(\mathbf{Z}_s)}\right] \\ &= \mathbb{E}\left[\frac{1}{\epsilon_{k-1}(\mathbf{Z}_{k-1})} \prod_{s=0}^{k-1} G_s^*(\mathbf{Z}_s)\right]. \end{aligned} \tag{21}$$

Combining (21) with inequality (20) we get that

$$\begin{aligned}&\mathbb{E}\Big[\prod_{s=0}^{k-1} G_s(\mathbf{Z}_s)\Big]\mathbb{E}\Big[\mathbb{E}[h(\mathbf{Z}_n)|\mathbf{Z}_k]^2 \prod_{s=0}^{k-1} G_s^{-1}(\mathbf{Z}_s)\Big] \\ &\geq \mathbb{E}\Big[\epsilon_{k-1}(\mathbf{Z}_{k-1}) \prod_{s=0}^{k-1} G_s^*(\mathbf{Z}_s)\Big]\mathbb{E}\left[\frac{1}{\epsilon_{k-1}(\mathbf{Z}_{k-1})} \prod_{s=0}^{k-1} G_s^*(\mathbf{Z}_s)\right] \end{aligned} \tag{22}$$

and using the Cauchy-Schwarz inequality on the right term, we get that

$$\begin{aligned}&\mathbb{E}\Big[\prod_{s=0}^{k-1} G_s(\mathbf{Z}_s)\Big]\mathbb{E}\Big[\mathbb{E}[h(\mathbf{Z}_n)|\mathbf{Z}_k]^2 \prod_{s=0}^{k-1} G_s^{-1}(\mathbf{Z}_s)\Big] \\ &\geq \mathbb{E}\Big[\prod_{s=0}^{k-1} G_s^*(\mathbf{Z}_s)\Big]^2 = \mathbb{E}\Big[\prod_{s=0}^{k-1} G_s^*(\mathbf{Z}_s)\Big]\mathbb{E}\left[\frac{\mathbb{E}\big[\mathbb{E}[h(\mathbf{Z}_n)|\mathbf{Z}_k]^2\big|\mathbf{Z}_{k-1}\big]}{\prod_{s=0}^{k-1} G_s^*(\mathbf{Z}_s)}\right]. \end{aligned} \tag{23}$$

By summing the inequalities (23) for each k, we easily see that

$$\sigma^2_{IPS,G} \geq \sigma^2_{IPS,G^*},$$

which completes the proof of the theorem. □

Remark that the optimal potential is unfortunately not positive everywhere, so it violates the hypothesis under which the TCL was proven in [8]. We claim that this question of the positiveness of the potential has not much interest in practice. Assume we take potential functions $(G_k)_{k<n}$ such that in the equation (20) we have $\epsilon_k(\mathbf{Z}_k) = 1$ and $\bar{\epsilon}_k(\mathbf{Z}_k) = \varepsilon > 0$ with ε very small. Choosing ε small enough, we can get $(G_k)_{k<n}$ as close has we want from $(G_k^*)_{k<n}$. With such potential functions $(G_k)_{k<n}$ and ε small enough, it is very likely that we would obtained the same samples in the algorithm as if we had taken the potentials $(G_k^*)_{k<n}$, and so we would have the same estimation.

Moreover, with such potential functions we would also have a TCL with a variance very close to σ^2_{IPS,G^*}. (According to (20) by choosing ε as small as we want, we get a variance as close as we want from σ^2_{IPS,G^*}.) In practice, positive potential functions $(G_k)_{k<n}$ with ε close to zero give the same results as the potentials $(G^*_k)_{k<n}$.

4. Application on a toy model

In this section we apply the IPS method on a toy system for which we have explicit formulas. The system under consideration is the Gaussian random walk $Z_{k+1} = Z_k + \varepsilon_{k+1}$, $Z_0 = 0$, where the $(\varepsilon_k)_{k\in\{1,\dots,n\}}$ are i.i.d. Gaussian random variables with mean zero and variance one. We explore two situations, one where the quantity to estimate,the optimal potential and the variance of the estimator can be calculated explicitly, and one where these quantities can be approximated by a large deviation inequality.

4.1. First example

In the first situation, taking $b, a > 0$ and $n \in \mathbb{N}\backslash\{0\}$, the goal is to compute the expectation $p_h = \mathbb{E}[h(\mathbf{Z}_n)]$ when $h(\mathbf{z}) = \exp[b(z-a)]$. As Z_n is a centered Gaussian of variance n, a simple calculation gives that for $k < n$:

$$\mathbb{E}[h(\mathbf{Z}_n)|Z_k = z] = \exp\left((n-k)\frac{b^2}{2} + b(z-a)\right). \tag{24}$$

Consequently we have that:

$$p_h = \exp\left(\frac{n}{2}b^2 - ab\right), \tag{25}$$

and that, for $k \geq 1$:

$$\sqrt{\frac{\mathbb{E}\Big[\mathbb{E}\big[h(\mathbf{Z}_n)\big|\mathbf{Z}_{k+1}\big]^2\big|\mathbf{Z}_k\Big]}{\mathbb{E}\Big[\mathbb{E}\big[h(\mathbf{Z}_n)\big|\mathbf{Z}_k\big]^2\big|\mathbf{Z}_{k-1}\Big]}} = \exp\left[-\frac{b^2}{2} + b(Z_k - Z_{k-1})\right] \tag{26}$$

$$\text{with} \qquad \mathbb{E}\Big[\mathbb{E}\big[\,h(\mathbf{Z}_n)\big|\mathbf{Z}_1\big]^2\big|\mathbf{Z}_0\Big] = \exp\left[\frac{(n+1)b^2}{2} + b(Z_0 - a)\right], \tag{27}$$

or equivalently that:

$$G^*_k(\mathbf{Z}_k) \propto \exp\big[b(Z_k - Z_{k-1})\big] \tag{28}$$

$$\text{with} \qquad G^*_0(\mathbf{Z}_0) \propto \exp\big[bZ_0\big]. \tag{29}$$

Using the equations (28) and (29) in (13), it can easily be shown that the variance of the IPS estimator with these optimal potential functions is:

$$\begin{aligned}\sigma^2_{IPS,G^*} &= n\left(\exp(b^2)-1\right)\exp\left[nb^2-2ab\right] \\ &= n\left(\exp(b^2)-1\right)p_h^2. \end{aligned} \tag{30}$$

It is notable that we obtain an asymptotically optimal variance [11], i.e. a variance proportional to p_h^2.

In order to confirm these theoretical results, we have carried out a simulation study. We have run the method 200 times with $N = 10^5$, $n = 10$ and different values of a and b, and for each of these values we have computed the mean of the estimation and the empirical variance of the estimation. The results are displayed in table 2, where we compare the theoretical value of p_h to the empirical mean of the 200 estimations, and σ^2_{IPS,G^*} to the empirical variance. As the theoretical values are close to the empirical ones, this confirms that the method is unbiased and that the variance given by the equation (13) is the right one. We also compare the empirical variance to the variance of the Monte-Carlo estimator, showing that, on this example, the IPS method provides a significant variance reduction with the optimal potential, as the variance is reduced by at least a factor 10^4 on the considered cases.

4.2. Second example

In the second situation, the goal is to compute the probability that Z_n exceeds a large positive value a. Therefore we take $h(\mathbf{Z}_n) = \mathbb{1}_{[a;+\infty)}(Z_n)$ so that

$$p_h = \mathbb{P}(Z_n \geq a). \tag{31}$$

In that case one can not compute $\mathbb{E}[h(Z_n)|Z_k = z]$ but the Chernov-Bernstein's inequality gives the following sharp exponential bound:

$$\mathbb{E}[h(\mathbf{Z}_n)|Z_k = z] \leq \exp\left[-\frac{(a-z)^2}{2(n-k)}\right], \tag{32}$$

from which we can deduce that:

$$\prod_{i=0}^{k-1} G_i^*(\mathbf{Z}_i) = \sqrt{\mathbb{E}\Big[\mathbb{E}\big[h(\mathbf{Z}_n)\big|\mathbf{Z}_k\big]^2\Big|\mathbf{Z}_{k-1}\Big]} \leq C_1 \exp\left[C_2 - \frac{(Z_{k-1}-a)^2}{2(n-k+2)}\right], \tag{33}$$

a	b	p_h	σ^2_{IPS,G^*}	σ^2_{MC}	$mean(\hat{p}_h)$	$\hat{\sigma}^2_{IPS,G^*}$
40	$\sqrt{log(1+\frac{1}{n})}$	6.98×10^{-6}	6.87×10^{-16}	6.98×10^{-6}	6.98×10^{-6}	6.72×10^{-16}
40	$\sqrt{log(1+\frac{2}{n})}$	9.51×10^{-8}	1.81×10^{-14}	9.51×10^{-8}	9.51×10^{-8}	1.76×10^{-14}
40	$\sqrt{log(1+\frac{3}{n})}$	4.69×10^{-9}	6.61×10^{-17}	4.69×10^{-9}	4.69×10^{-9}	7.37×10^{-17}
40	$\sqrt{log(1+\frac{4}{n})}$	4.50×10^{-10}	8.13×10^{-19}	4.50×10^{-10}	4.50×10^{-10}	7.39×10^{-17}
35	$\sqrt{log(1+\frac{1}{n})}$	3.27×10^{-5}	1.09×10^{-9}	3.27×10^{-5}	3.27×10^{-5}	9.69×10^{-10}
35	$\sqrt{log(1+\frac{2}{n})}$	8.04×10^{-7}	1.29×10^{-12}	8.04×10^{-7}	8.04×10^{-7}	1.35×10^{-12}
35	$\sqrt{log(1+\frac{3}{n})}$	6.07×10^{-8}	1.11×10^{-14}	6.07×10^{-8}	6.07×10^{-8}	1.06×10^{-14}
35	$\sqrt{log(1+\frac{4}{n})}$	8.19×10^{-9}	2.69×10^{-16}	8.19×10^{-9}	8.19×10^{-9}	2.85×10^{-16}

TABLE 2: Theoretical and empirical comparisons (example 1)

$N = 2 * 10^5, n = 10$

where C_1 and C_2 are some constants independent of $\mathbf{Z}_{k-1}$. One can therefore try to set the product of the potentials equal to this upper bound, which yields:

$$\text{for } k \geq 1, \qquad G_k(\mathbf{Z}_k) \propto \exp\left[-\frac{(Z_k - a)^2}{2(n-k+1)} + \frac{(Z_{k-1} - a)^2}{2(n-k+2)}\right] \tag{34}$$

$$\text{and} \qquad G_0(\mathbf{Z}_0) \propto \exp\left[-\frac{(Z_0 - a)^2}{2(n-1)}\right]. \tag{35}$$

Similarly as for the first example, we have carried out a simulation study. We have run the method 200 times with the potentials defined in equation (34) and (35) and taking $N = 2 \times 10^5$, and $n = 10$, and different values of a, and for each of these values we have computed the empirical mean of the estimation and the empirical variance of the estimation. The results are displayed in table 3. We compare these estimations with the actual values of p_h and the variance of the Monte-Carlo method σ^2_{MC}, showing that the potentials built with the Chernov-Bernstein large deviation inequality and our formula yield a significant variance reduction. Indeed the variance reduction compared to the Monte-Carlo method is at least by a factor 8500, and at best by a factor $2.4 * 10^6$. We also compared the efficiency of different potentials.

a	p_h	σ^2_{IPS,G^*}	σ^2_{MC}	$mean(\hat{p}_h)$	$\hat{\sigma}^2_{IPS,G}$
$4\sqrt{n}$	$3.17 * 10^{-5}$	?	$3.17 * 10^{-5}$	$3.18 * 10^{-5}$	$8.14 * 10^{-8}$
$5\sqrt{n}$	$2.87 * 10^{-7}$	?	$2.87 * 10^{-7}$	$2.86 * 10^{-7}$	$1.88 * 10^{-11}$
$6\sqrt{n}$	$9.87 * 10^{-10}$	?	$9.87 * 10^{-10}$	$9.67 * 10^{-10}$	$6.63 * 10^{-16}$
$7\sqrt{n}$	$1.28 * 10^{-12}$	?	$1.28 * 10^{-12}$	$1.29 * 10^{-12}$	$4.47 * 10^{-21}$

TABLE 3: Theoretical and empirical comparisons (example 2)
results obtained with $N = 10^5$ and $n = 10$

We run the method with 1) the potential used on a Gaussian random walk in [13]: $G_k(\mathbf{Z}_k) = \exp[\alpha(Z_k - Z_{k-1})]$ where the parameter was optimized to $\alpha = 1.1$, 2) the potential built with the Chernov-Bernstein large deviation inequality, and 3) with the optimal potential that we computed using Gaussian quadrature formulas. The results are displayed in table 4, and show that indeed the potential functions $(G^*_k)_{k<n}$, where $G^*_k(\mathbf{Z}_k) = \sqrt{\frac{\int_{\mathbb{R}}\left(\int_a^\infty \exp\left[-\frac{(z_n - z_{k+1})^2}{2(n-k+1)}\right] dz_n\right)^2 \exp\left[-\frac{(z_{k+1} - Z_k)^2}{2}\right] dz_{k+1}}{\int_{\mathbb{R}}\left(\int_a^\infty \exp\left[-\frac{(z_n - z_k)^2}{2(n-k)}\right] dz_n\right)^2 \exp\left[-\frac{(z_k - Z_{k-1})^2}{2}\right] dz_k}}$, yield the best variance.

$G_k(\mathbf{Z}_k)$	$mean(\hat{p}_h)$	$\hat{\sigma}^2_{IPS,G}$
$\exp\left[\alpha(Z_k\text{-}Z_{k\text{-}1})\right]$	$1.04 * 10^{-6}$	$2.44 * 10^{-10}$
$\exp\left[-\frac{(Z_k\text{-}a)^2}{2(n\text{-}k+1)} + \frac{(Z_{k\text{-}1}\text{-}a)^2}{2(n\text{-}k+2)}\right]$	$1.03 * 10^{-6}$	$1.78 * 10^{-10}$
$G^*_k(\mathbf{Z}_k)$	$1.04 * 10^{-6}$	$1.62 * 10^{-10}$

TABLE 4: Comparisons of the efficiency of potentials (example 2)
results obtained for $p_h = 1.05 * 10^{-6}$ $N = 2000$, $n = 10$, $a = 15$, $\alpha = 1.1$

5. Discussion and implication

In this paper we give a closed form expression of the optimal potential functions for the IPS method with multinomial resampling, and for its minimal variance. The existence of optimal potential function, proves that the possible variance reduction of an IPS method is lower-bounded. The expressions have been validated analytically, and the expression for the minimal variance has been empirically confirmed in toy examples.

Furthermore the results found in the literature seem consistent with our findings. Indeed, in [10] the authors made the observation that it seemed better to build a potential which increments of an energy function, this observation is confirmed as the optimal potential is the multiplicative increment of the quantity $\sqrt{\mathbb{E}\Big[\mathbb{E}\big[h(\mathbf{Z}_n)\big|\mathbf{Z}_k\big]^2\big|\mathbf{Z}_{k-1}\Big]}$ which is then the optimal energy function. Also, the fact that in [15] the authors find better results with time-dependent potential is explained by the fact the expression of the optimal potential shows a dependency on k. Finally, as splitting methods can be viewed as a version of the IPS-method with indicator potential functions, our results show that the selections of splitting algorithms are not optimal, and could be improved by using information on the expectations $\mathbb{E}\big[h(\mathbf{Z}_n)\big|\mathbf{Z}_k = \mathbf{z}\big]$.

The optimal potential functions may be hard to find in practice. Indeed, the expectations $\mathbb{E}\big[h(\mathbf{Z}_n)\big|\mathbf{Z}_k = \mathbf{z}\big]$ play a big role in the expression of the optimal potentials, but if we are trying to assess $p_h = \mathbb{E}\big[h(\mathbf{Z}_n)\big]$, we typically lack information about the expectations $\mathbb{E}\big[h(\mathbf{Z}_n)\big|\mathbf{Z}_k = \mathbf{z}\big]$. If no information on the expectation $\mathbb{E}\big[h(\mathbf{Z}_n)\big|\mathbf{Z}_k = \mathbf{z}\big]$ is available, it might be preferable to use more naive variance reduction method, where no input functions are needed. In such context, the Weighted Ensemble (WE) method

[2, 1] seems to be a good candidate, as it does not take in input potential functions but only a partition of the state space. Conversely if the practitioner has information about the expectation $\mathbb{E}\big[h(\mathbf{Z}_n)\big|\mathbf{Z}_k = z\big]$, this information could be used to derive very efficient potentials.

The knowledge of these expectations is therefore crucial for a well optimized use of the IPS method, but it is interesting to remark that the same knowledge seem to be crucial for a well optimized importance sampling method. Indeed the optimal density of importance sampling (that gives a zero variance) depends on $\mathbb{E}\big[h(\mathbf{Z}_n)\big|Z_{[0,t]} = z_{[0,t]}\big]$ when it is used on a piecewise deterministic process [7]. This confirms the well known fact that, with a good knowledge of the dynamic of the process $(\mathbf{Z}_k)_{k\geq 0}$, the importance sampling methods is preferable to the IPS. Nonetheless the IPS method may still be preferred to the importance sampling methods when that knowledge is not available. When one fears to be in an over-biasing situation the IPS may be preferable, as in the IPS method we do not alter the propagation, the over-biasing phenomenon should then be less important than with importance sampling methods.

References

[1] D. Aristoff. Optimizing weighted ensemble sampling for steady states and rare events. *ArXiv e-prints*, (arXiv1806.00860), June 2018.

[2] David Aristoff. Analysis and optimization of weighted ensemble sampling. *ESAIM: Mathematical Modelling and Numerical Analysis*, 52(4):1219–1238, 2018.

[3] F. Cérou, P. Del Moral, A. Guyader, et al. A nonasymptotic theorem for unnormalized feynman–kac particle models. In *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques*, volume 47, pages 629–649. Institut Henri Poincaré, 2011.

[4] F. Cérou and A. Guyader. Adaptive multilevel splitting for rare event analysis. *Stochastic Analysis and Applications*, **25**(2):417–443, 2007.

[5] F. Cérou, F. LeGland, P. Del Moral, and P. Lezaud. Limit theorems for the multilevel splitting algorithm in the simulation of rare events. In *Simulation Conference, Orlando, United States. Proceedings of the Winter*, pages 10–pp. IEEE, 2005.

[6] Hock Peng Chan, Tze Leung Lai, et al. A general theory of particle filters in hidden markov models and some applications. *The Annals of Statistics*, 41(6):2877–2904, 2013.

[7] H. Chraibi, A. Dutfoy, T. Galtier, and J. Garnier. Speed-up reliability assessment for multi-component systems: importance sampling adapted to piecewise deterministic markovian processes. ESREL conference, 2016.

[8] P. Del Moral. In *Feynman-Kac Formulae, Genealogical and Interactig Particle Systems With Applications*. Springer, 2004.

[9] P. Del Moral, A. Doucet, and A. Jasra. Sequential monte-carlo samplers. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **68**(3):411–436, 2006.

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

[11] P. Dupuis and H. Wang. Importance sampling, large deviations, and differential games. *Stochastics and Stochastic Reports*, 76(6):481–508, 2004.

[12] M. Balesdent J. Marzat J. Morio, D. Jacquemart. Optimisation of interacting particle systems for rare event estimation. *Computational Statistics & Data Analysis*, **66**:117–128, 2013.

[13] D. Jacquemart and J. Morio. Tuning of adaptive interacting particle system for rare event probability estimation. *Simulation Modelling Practice and Theory*, **66**:36–49, 2016.

[14] A. Lee and N. Whiteley. Variance estimation in the particle filter. *Biometrika*, 105(3):609–625, 2018.

[15] J. Wouters and F. Bouchet. Rare event computation in deterministic chaotic systems using genealogical particle analysis. *Journal of Physics A: Mathematical and Theoretical*, **49**(37):374002, 2016.