manuscript No.
(will be inserted by the editor)

Supplementary material for the manuscript "Analysis of adaptive directional stratification for the controlled estimation of rare event probabilities"

Miguel Munoz Zuniga · Josselin Garnier · Emmanuel Remy · Etienne de Rocquigny

Received: date / Accepted: date

Abstract In the context of structural reliability, a small probability to be assessed, a high computational time model and a relatively large input dimension are typical constraints which brought together lead to an interesting challenge. Indeed, in this framework many existing stochastic methods fail in estimating the failure probability with robustness.
Therefore, the aim of this article is to present and prove theoretical results about the validity of an original method we have introduced to overcome the specific constraints mentioned above. This new method turns out to be very competitive compared with the existing techniques. It is a variant of accelerated Monte-Carlo simulation method, named ADS-2 - Adaptive Directional Stratification. It combines, in a two steps adaptive strategy, the stratification into quadrants and the directional simulation techniques. Two ADS-2 estimators are presented and their properties are studied.

Keywords Monte-Carlo · Simulation · Directional · Stratification · Adaptive · Small probabilities

M. Munoz Zuniga
Laboratoire de Probabilités et Modèles Aléatoires, Université Paris VII, Case courrier 7012, 2 place Jussieu, 75251 Paris cedex 05, France
Tel.: +33630639428
E-mail: mmunozzu@hotmail.com

J. Garnier
Laboratoire de Probabilités et Modèles Aléatoires, Université Paris VII, Case courrier 7012, 2 place Jussieu, 75251 Paris cedex 05, France

E. Remy
EDF R&D, 6 quai Watier, 78400 Chatou, France

E. de Rocquigny
Ecole Centrale de Paris, Grande Voie des Vignes, 92295 Chatenay-Malabry, France

1 Introduction

Our work has been motivated by the continuing difficulties encountered in some key industrial applications when tackling the issue of controlling the robustness of existing algorithms. Two different models have been considered: a flood model and a nuclear reactor pressurized vessel fracture mechanics model (Munoz Zuniga et al. 2008 and 2010 [29,30]). The nuclear reactor pressurized vessel case study is the main motivation of this work: the mathematical and numerical constraints inherent to this model constituted our framework. Let us present this framework and the general problem we are dealing with.
One way to assess the reliability of a structure from physical considerations is to use a probabilistic approach: it includes the joint probability distribution of the random input variables of a numerical deterministic model representing the physical behavior of the studied structure, for instance its fail-

ure margin. We consider a real-valued failure function, $G : \mathbb{D} \subset \mathbb{R}^p \to \mathbb{R}$, whose input vector $\mathbf{X}$ is random. Then, we assume that the probability density function of the random vector $\mathbf{X}$ exists and is known.

In this context, we want to assess the failure probability P_f:

$$P_f = \mathbb{P}(G(\mathbf{X}) < 0) = \int_{G(\mathbf{x})<0} f(\mathbf{x})d\mathbf{x} \tag{1}$$

with G the failure function defined over $\mathbb{D}$, $\mathbf{X} = (X^1, ..., X^p)$ the p-dimensional random vector of the input variables and f the density function of $\mathbf{X}$. We partition the set where G is defined, $\mathbb{D}$, into three subsets: the reliability (or safe) domain, denoted by $\mathbb{D}_r := \{\mathbf{x} \in \mathbb{D}, G(\mathbf{x}) > 0\}$, the failure domain, denoted by $\mathbb{D}_f := \{\mathbf{x} \in \mathbb{D}, G(\mathbf{x}) < 0\}$, and the boundary between the two previous subsets, the limit state surface, denoted by $\mathbb{D}_l := \{\mathbf{x} \in \mathbb{D}, G(\mathbf{x}) = 0\}$. Furthermore, four key features characterize our agenda.

- (C_0) : G may be complex and greedy in computational resources: even when involving high performance computing, industrial constraints limit the number of evaluations of G to a few thousands.
- (C_1) : No continuity or derivability assumptions are considered for G.
- (C_2) : The failure is a rare event, which means that P_f is very small. In this work, we will consider that a small probability is a probability lower than 10^{-3}.
- (C_3) : The results must be robust, i.e. with explicit and trustworthy error control.

The first three constraints correspond to our working hypotheses and the last constraint is the key goal motivating this research. Indeed, small failure probability to be assessed, robustness of the estimation, costly computational time model and relatively large input dimension are typical constraints which brought together lead to a real challenge. A well-developed state of the art of stochastic algorithms for the failure probability estimation can be found in Rubinstein and Kroese (2007) [34], Gille-Genest (1999) [16] and Cannamela (2007) [5] with a large bibliography. For instance, the numerical integration methods represent a possible solution. Nevertheless, they are not effective when the dimension p of $\mathbb{D}$ is large. Although we can notice that methods such as sparse grid are currently developed to overcome the curse of dimensionality, all of them require some assumptions on the failure function to be integrated; one can see Dirnstorfer and Bungartz (1985) [4], Griebel and Gerstner (1998) [14] and Griebel et al. (2003) [15]. Another standard solution consists in keeping the real physical model coupled with the FORM/SORM numerical approximation as in Lind et al. (2000) [27] and Ditlevsen et al.(1996) [26]. However, in non-analytic cases, the error carried out is not easily assessable, even if the calculation times remain reasonable (Lind et al. 2000 [27]): thus the robustness of this kind of method is not ensured. The last solution is to evaluate the failure probability (1) by Monte-Carlo simulation. This method is reliable and controllable thanks to probabilistic theorems, which give an accurate statistical measure of the estimation error in the form of confidence intervals. Large failure probabilities are mostly satisfactorily treated by crude Monte-Carlo sampling as presented in Rubinstein et al. (2007) [34], Helton (1993) [17] and Lapeyre et al. (1997) [21]. Nevertheless, the required number of simulations is unrealistic for a rare failure problem and a complex physical model. There exist accelerated methods which reduce the variance of the failure probability estimator but they do not fulfill to all the constraints we assume in this article, particularly (C_2) (see Siegmund 1976 [36], Cochran 1977 [7], Ripley 1987 [33], Zhang 1996 [40], Johnson et al. 2005 [18], Kroese and Rubinstein 2007 [34]). Another class of accelerated Monte-Carlo methods have been specifically developed to estimate accurately small probabilities with a large input dimension. The primary approaches are splitting and importance sampling. They often rely on Markov chains simulations, see for instance: Au et al. 2001 [1], Rubinstein et al. 2002 [19], Garnier et al. 2005 [10], L'Ecuyer et al. 2006 [22], Demers et al. 2007 [23], Dean et al. 2009 [9], Lagnoux-Renaudie 2009 [20], Kroese et al. 2009 [6]. These methods are very robust with respect to the dimension. However, the minimal number of simulations require to obtain a reasonable convergence is still to large for our framework (C_0). In order to illustrate these limitation, numerical results are provided in section 5. All the previous methods may be combined with a response surface approach, replacing the initial physical model by a simplified one with numerical approximation. This alternative includes for instance the largely-developing non intrusive chaos polynomial approaches (Blatman et al. 2011 and 2010 [3,2], Le maître et al. 2009 [8], Sudret 2008 [38], Todor et al. 2007 [39], Rabitz et al. 2006 [24], Soize et al. 2004 [37]) and the recently popular Kriging methods (Fang et al. 2006 [12], Rasmussen et al. 2006 [32], Santner et al. 1999 [35]). However, the control of the errors carried out by these methods, generally due to the limited-order to which the model approximation is undertaken, is sometimes not available (computing error bounds typically requires higher-order developments), and when available, the errors must be added to the unavoidable error brought by the estimation. In this paper, we choose to focus on the simulation methods rather than the response surface techniques. Within the context of structural reliability applications, the Monte-Carlo sampling basis remains generally the reference. Here, we consider the accelerated Monte-Carlo methods and try to develop a specific one, which "converges" as fast as possible and enables to obtain an estimation error control in a reasonable number of simulations.

2 Preliminary: stratification, directional sampling and optimization

The idea is to take advantage of the possibilities offered by the stratification and directional simulation methods: optimal allocation result, adaptive strategy, efficient small probability estimation and reasonable calculation time.

Let us assume that we have a p-dimensional random input vector $\mathbf{X}$ and we want to estimate $P_f = \mathbb{E}\big(\mathbb{1}_{G(\mathbf{X})<0}\big)$. We first move the problem into a "spherical space": the Gaussian space, using the Nataf transformation, $\mathbf{U} = T_N(\mathbf{X})$ (Kiureghian and Liu 1986 [25]). In other words, the expectation estimation becomes:

$$P_f = \mathbb{E}\big(\mathbb{1}_{G\circ T_N^{-1}(\mathbf{U})<0}\big) = \mathbb{E}\big(\mathbb{1}_{g(\mathbf{U})<0}\big) \tag{2}$$

with $g = G \circ T_N^{-1}$ and $\mathbf{U}$ a random vector with Gaussian and independent components.

Then, $\mathbf{U}$ can be decomposed into the product $R\mathbf{A}$ (R and $\mathbf{A}$ independent) with R^2 a chi-square random variable and $\mathbf{A}$ a random vector uniformly distributed over the unit sphere $S_{p-1} \subset \mathbb{R}^p$ (Fang et al. 1990 [13], Gille-Genest 1999 [16]). We will denote $\mathcal{L}(\mathbf{A})$ the distribution of $\mathbf{A}$. Then, Denoting f_R the probability density function of R and using the conditional expectation properties, we have:

$$\begin{aligned} P_f &= \mathbb{E}\big(\mathbb{1}_{g(R\mathbf{A})<0}\big) = \mathbb{E}\Big[\mathbb{E}\big(\mathbb{1}_{g(R\mathbf{A})<0}|\mathbf{A}\big)\Big] \\ &= \mathbb{E}\big(\xi(\mathbf{A})\big) \end{aligned} \tag{3}$$

with $\xi(\mathbf{a}) = \int \mathbb{1}_{g(r\mathbf{a})<0} f_R(r)dr$ a bounded function.

The directional sampling method is based on the fact that we are able to calculate $\xi(\mathbf{a})$ for any $\mathbf{a}$. Indeed, the conditional expectation $\xi(\mathbf{a})$, which is the probability: $\mathbb{P}(g(R\mathbf{a}) < 0)$, is then given by:

$$\xi(\mathbf{a}) = \sum_{i=1}^{s} \chi_p^2\big(r_{u,i}^2(\mathbf{a})\big) - \chi_p^2\big(r_{l,i}^2(\mathbf{a})\big) \tag{4}$$

with χ_p^2 the chi-square cumulative distribution function (cdf), $r_{u,i}(\mathbf{a})$ and $r_{l,i}(\mathbf{a})$ respectively the upper and the lower bounds of the i-th interval where $g(r\mathbf{a}) < 0$ and s the number of intervals (see Ditlevsen and Madsen 1996 [26]). $r_{u,i}(\mathbf{a})$ and $r_{l,i}(\mathbf{a})$ are approximated thanks to root-finding algorithms as the dichotomic, the secant, the inverse quadratic interpolation or the Brent methods. A stop criterion for the dichotomic algorithm in the directional sampling method, applied to the estimation of small failure probability, can be found in [28]. This latter is based on the control of the variance of the estimator which increases with the add of the dichotomic algorithm intrinsic errors. The directional simulation estimator reduces the variance in comparison with the standard Monte-Carlo sampling method, as can be shown using Jensen's inequality (see Ditlevsen and Madsen 1996 [26]). At this point, we introduce the stratification method by

partitioning the "directional space", in other words the space where $\mathbf{A}$ takes its values, that is to say the $(p-1)$-dimensional unit sphere $S_{p-1} \subset \mathbb{R}^p$, into m strata and we denote $I := \{1, ..., m\}$. The natural strata adapted to directional draws are cones and partitioning S_{p-1} is equivalent to make a partition of the general space into cones. Let us denote by $(q_i)_{i\in I}$ a partition of S_{p-1} into m strata. Let us denote by n the total number of directional draws, e.g. the number of directions we want to simulate. Let us consider an allocation of the n directional draws in each stratum described by the sequence $\mathbf{w} = (w_i)_{i\in I}$ such that $\sum_{i=1}^m w_i = 1$. The number of draws in the i-th stratum is $n_i = \lfloor n w_i \rfloor$ with $\lfloor . \rfloor$ the floor function (we neglect the rounding errors in the analysis).
We can express the expectation P_f as:

$$P_f = \sum_{i=1}^m \mathbb{P}(\mathbf{A} \in q_i)\mathbb{E}\big(\xi(\mathbf{A}^i)\big) = \sum_{i=1}^m \rho_i P_i \tag{5}$$

with $\mathbf{A}^i \sim \mathcal{L}(\mathbf{A}|\mathbf{A} \in q_i)$ and $P_i = \mathbb{E}\big(\xi(\mathbf{A}^i)\big)$.
Now, we estimate P_i by drawing n_i directions in the i-th stratum. This can be done by using a simple rejection method. We get:

$$\hat{P}_i = \frac{1}{n_i}\sum_{j=1}^{n_i} \xi(\mathbf{A}_j^i) \tag{6}$$

with $(\mathbf{A}_j^i)_{j=1,\dots,n_i}$ a family of independent and identically distributed (i.i.d.) random vectors with the distribution $\mathcal{L}(\mathbf{A}|\mathbf{A} \in q_i)$.
We obtain the following unbiased estimator by coupling the Directional Sampling and the Stratification methods:

$$\hat{P}_f^{DSS} = \sum_{i=1}^m \rho_i \hat{P}_i \tag{7}$$

and its variance is given by σ_{DSS}^2/n with

$$\sigma_{DSS}^2 = \sum_{i=1}^m \frac{\rho_i^2 \sigma_i^2}{w_i} \tag{8}$$

where $\sigma_i^2 = Var\big(\xi(\mathbf{A}^i)\big)$.
Then we get the following asymptotic result which gives the rate of convergence of the DSS estimator (coupling stratification and directional simulation) for large values of n. The proof can be found in section 7.1.

Proposition 1 *The DSS estimator is consistent and*

$$\sqrt{n}(\hat{P}_f^{DSS} - P_f) \xrightarrow[n\to+\infty]{\mathcal{L}} \mathcal{N}(0, \sigma_{DSS}^2).$$

Now, just like in the standard stratification method, we can determine the optimal distribution of directional draws in each stratum in order to minimize the variance estimator given by (8), which can be useful for the design of an adaptive strategy. From now on, we assume that the following assumption is satisfied:

$(H)\ \ \exists i \in I,\ \ \sigma_i > 0.$

We need to solve the following problem:

$$(E_1) \left\{ \begin{array}{l} \min\limits_{\mathbf{w}\in\mathbb{R}_+^m} \left(\frac{1}{n}\sum\limits_{i=1}^m \frac{\rho_i^2\sigma_i^2}{w_i}\right) \\ \text{under the constraint: } \sum\limits_{i=1}^m w_i = 1. \end{array}\right. \tag{9}$$

The unique solution is given by the following proposition which is demonstrated in Cochran 1977 [7]. The proof is rather standard but we give it for consistency in section 7.2.

Proposition 2 *The optimization problem* (E_1) *has for unique solution the sequence*

$$\mathbf{w_{opt_1}} := (w_i^{opt_1})_{i\in I}$$

defined for all $i \in I$ *by*

$$w_i^{opt_1} = \frac{\rho_i\sigma_i}{\sum\limits_{j=1}^m \rho_j\sigma_j}$$

and the optimal variance obtained with this allocation is: $\sigma_{opt_1}^2/n$ *with*

$$\sigma_{opt_1}^2 := \left(\sum_{i=1}^m \rho_i\sigma_i\right)^2.$$

We observe that the optimal allocation in a stratum is proportional to the variance in this stratum. That means that we must give large weights to the strata with the largest uncertainties, i.e. variances.
Now, we can see that the variance of the estimator $\hat{P}_f^{DSS}$ with the optimal allocation is smaller than the one with the proportional one. Besides, we can demonstrate that the variance of $\hat{P}_f^{DSS}$ with the

proportional allocation, and a fortiori with the optimal one, is: (a) smaller than the variance of the usual stratified estimator (in the Gaussian space) with a proportional or an optimal allocation, (b) smaller than the variance of the standard directional simulation method. All these variances are also smaller than the one of the standard Monte-Carlo. We can notice that the variance reduction is not ensured with an arbitrary allocation: if the allocation is strongly inappropriate then the variance can increase in comparison with the MC method.
Just before starting the presentation of an adaptive strategy, we can imagine that in an iterative approach, we will certainly have to draw directions in a stratum where we have already drawn some directions. Let us consider that the iterative approach consists of 2 steps. We split the total number of simulations n into two parts $\gamma_1(n)n$ and $\gamma_2(n)n$ such as $\gamma_1(n) + \gamma_2(n) = 1$. Let us assume that in the first step, we have already drawn a sizeable proportion, β_i, of $\gamma_1(n)n$ in the i-th stratum and that we want now to draw a proportion w_i of $\gamma_2(n)n$ in this stratum. At the end, we will have drawn a proportion $\gamma_1(n)\beta_i n + \gamma_2(n)w_i n$ in the i-th stratum, for all i. Of course $\sum_{i=1}^m \beta_i = 1$ and $\sum_{i=1}^m w_i = 1$. Then, the variance of the estimator will be:

$$\frac{1}{n}\sum_{i=1}^{m}\frac{\rho_i^2\sigma_i^2}{\gamma_1(n)\beta_i + \gamma_2(n)w_i}. \tag{10}$$

We can now try to calculate the allocation $(w_i)_{i\in I}$ which will minimize this variance, by solving the following optimization problem:

$$(E_2)\begin{cases} \min_{w\in\mathbb{R}_+^m}\left(\frac{1}{n}\sum_{i=1}^{m}\frac{\rho_i^2\sigma_i^2}{\gamma_1(n)\beta_i + \gamma_2(n)w_i}\right) \\ \text{under the constraint: } \sum_{i=1}^{m} w_i = 1 \end{cases}, \tag{11}$$

in which the β_j are given. Let us suppose that all $\sigma_i \neq 0$. If not, we can rewrite the problem only with the σ_i different from zero and for the σ_i that are null we take $w_i = 0$.
There exists a unique solution for the problem (E_2): it can be determined by the following algorithm.

Algorithm 1

(a) Compute, for $i \in I$, the quantities $\frac{\beta_i}{\rho_i\sigma_i}$ and sort them in decreasing order, denoting them $\frac{\beta_{(i)}}{\rho_{(i)}\sigma_{(i)}}$. Compute, for $i = 1, ..., m-1$, the quantities

$$\frac{1 + \frac{\gamma_1(n)}{\gamma_2(n)}\sum_{j=i+1}^{m}\beta_{(j)}}{\frac{\gamma_1(n)}{\gamma_2(n)}\sum_{j=i+1}^{m}\rho_{(j)}\sigma_{(j)}}.$$

(b) Denote by i^* the largest i such that

$$\frac{\beta_{(i)}}{\rho_{(i)}\sigma_{(i)}} \geq \frac{1 + \frac{\gamma_1(n)}{\gamma_2(n)}\sum_{j=i+1}^{m}\beta_{(j)}}{\frac{\gamma_1(n)}{\gamma_2(n)}\sum_{j=i+1}^{m}\rho_{(j)}\sigma_{(j)}}.$$

If this inequality is false for all i, then, by convention, $i^* = 0$.

(c) Finally, for $i \leq i^*$, set $w_{(i)} = 0$, and for $i > i^*$, set:

$$w_{(i)} = \frac{\rho_{(i)}\sigma_{(i)}}{\sum_{j=i^*+1}^{m}\rho_{(j)}\sigma_{(j)}}\left(1 + \frac{\gamma_1(n)}{\gamma_2(n)}\sum_{j=i^*+1}^{m}\beta_{(j)}\right) - \frac{\gamma_1(n)}{\gamma_2(n)}\beta_{(i)}$$

and turn back to the initial indexation.

Let us denote by $K \subset I$ the subset defined by:

$$K = \left\{(i) \in I, i \leq i^*\right\} \tag{12}$$

and by K^c the complementary of K. The subset K represents the set of indices corresponding to the strata which do not need extra draws than the ones made in the first step, and K^c the one where supplementary draws are needed. The following proposition is an extension of Proposition 1 and its proof can be found in section 7.3. It gives the general optimal allocation expression and the associated minimal variance.

Proposition 3 *The optimization problem* (E_2) *has for unique solution the sequence*

$$\mathbf{w_{opt_2}} := (w_i^{opt_2})_{i\in I}$$

defined for all $i \in I$ *by*

$$w_i^{opt_2} = 0 \quad \text{if } i \in K$$

and

$$w_i^{opt_2} = \frac{\rho_i\sigma_i}{\sum_{j\in K^c}\rho_j\sigma_j}\left(1 + \frac{\gamma_1(n)}{\gamma_2(n)}\sum_{j\in K^c}\beta_j\right) - \frac{\gamma_1(n)\beta_i}{\gamma_2(n)} \quad \text{if } i \in K^c$$

with (K, K^c) *the partition of* I *defined by algorithm 1.*
The optimal variance obtained with this allocation is: $\sigma^2_{opt_2}/n$ *with*

$$\sigma^2_{opt_2} := \sum_{i\in K}\frac{\rho_i^2\sigma_i^2}{\gamma_1(n)\beta_i} + \frac{1}{\gamma_2(n) + \gamma_1(n)\sum_{i\in K^c}\beta_j} \times \left(\sum_{i\in K^c}\rho_i\sigma_i\right)^2.$$

Thus, from proposition 3, we deduce that for $i \in K^c$, i.e. for a stratum where additional draws must be performed, the number of directional draws that must be achieved in it with the $\gamma_2(n)n$ remaining directional draws is:

$$\gamma_2(n)nw_i^{opt_2} = \frac{\rho_i\sigma_i}{\sum_{j\in K^c}\rho_j\sigma_j}\Big(\gamma_2(n)n + \gamma_1(n)n\sum_{j\in K^c}\beta_j\Big) - \gamma_1(n)n\beta_i. \tag{13}$$

So, we need to perform an optimal allocation, proportional to the variance of the considered stratum, over the number of draws of the second step plus the already allocated number of draws in the first step and subtract the number of directional draws already made in the considered stratum: $\gamma_1(n)n\beta_i$.
Also, we can compare the optimal variances we obtained, on the one hand, with allocation $\mathbf{w_{opt_1}}$ given by solving (E_1) and, on the other hand, with allocation $\mathbf{w_{opt_2}}$ given by solving (E_2). It seems clear that we will reach a minimal variance when the optimal allocation is determined over the total number of draws as in the problem (E_1). Indeed in the other case, (E_2), the first set of draws is "lost" in the sense that they are not necessarily allocated in the optimal way. This fact is confirmed by the following result proved in section 7.4.

Proposition 4 *The optimal variances* $\sigma^2_{opt_1}$ *and* $\sigma^2_{opt_2}$*, respectively reached with the optimal allocation* $\mathbf{w_{opt_1}}$ *and* $\mathbf{w_{opt_2}}$*, verify:*

$$\sigma^2_{opt_1} \le \sigma^2_{opt_2}.$$

There is equality $\sigma^2_{opt_1} = \sigma^2_{opt_2}$ when $K^c = I$ i.e. when, in the second step, all the strata require additional draws.

3 ADS-2 method

3.1 General principle of the ADS-2 method

Now, we have all the tools to set up our 2-steps adaptive method with an estimator coupling stratification and directional sampling. We will call this estimator the 2-adaptive directional stratification (ADS-2) estimator. It will be denoted by $\hat{P}_f^{ADS-2}$.
Let us explain our strategy. We split the total number n of draws into two parts: $\gamma_1(n)n$ and $\gamma_2(n)n$ such that $\sum_{i=1}^2\gamma_i(n) = 1$ and $(\gamma_i(n))_{i=1,2}$ are functions of variable n. The first part, $\gamma_1(n)n$, will be used to estimate the optimal allocation minimizing the variance (learning step), and the second part, $\gamma_2(n)n$, will be used to estimate the failure probability (estimation step), using the estimation of the optimal allocation.
For instance, we can take $\gamma_1(n) = n^{l-1}$ with $l \in (0,1)$, or $\gamma_1(n) = \gamma_1$ with $\gamma_1 \in (0,1)$ and, in an obvious way, we have $\gamma_2(n) = 1 - \gamma_1(n)$. Let us set the strata as the quadrants of the Gaussian space, denoted by $(q_i)_{i\in I}$ with $m = 2^p$ the number of quadrants. Indeed, without any other information, the most natural choice is quadrants, which are well adapted for directional draws and enable a good survey of the Gaussian space. Also, without prior information, we should choose a prior uniform allocation. We denote it $(w_i)_{i\in I}$, and as we considered quadrants: $w_i = \rho_i = 1/m$ for all $i \in I$. Our method can be readily extended

to the case in which a prior information is available on the optimal allocation. From now on, $n_i = \lfloor \gamma_1(n) n w_i \rfloor$ for all $i \in I$ stands for the number of draws achieved in each stratum in the estimation step. Let us suppose that we use the directional and stratification estimator $\hat{P}_f^{DSS}$. Consequently, the idea is to replace the initial allocation by an estimation of the optimal allocation made with the $\gamma_1(n)n$ first draws: optimal allocation $\mathbf{w_{opt_1}}$ and $\mathbf{w_{opt_2}}$, respectively defined in propositions 2 and 3, are the ones on which we will refer.
But first, we need to define our estimators. Two possibilities can be considered: we can choose either a non-recycling estimator using an estimation of $\mathbf{w_{opt_1}}$ or a recycling estimator using an estimation of $\mathbf{w_{opt_2}}$.

3.2 The recycling ADS-2 estimator: definition and properties

The recycling ADS-2 estimator uses the $\gamma_1(n)n$ first draws, that are used for the optimal allocation estimation step, for the expectation estimation:

$$\hat{P}_{f,r}^{ADS-2} = \sum_{i=1}^{m} \rho_i \frac{1}{N_i^r} \sum_{j=1}^{N_i^r} \xi(\mathbf{A}_j^i) \tag{14}$$

with

$$N_i^r := \lfloor (1-\epsilon(n))\big(n_i + n_i^r\big) + \epsilon(n)\rho_i n \rfloor, \tag{15}$$

where $n_i^r = \left\lfloor \gamma_2(n) n \tilde{W}_i^r \right\rfloor$ and $(\tilde{W}_i^r)_{i\in I}$ are given by proposition 3 and algorithm 1, substituting into them the deterministic sequence $(\sigma_i)_{i\in I}$ by an estimation of it, $(\tilde{\sigma}_i)_{i\in I}$, given, for all $i \in I$, by:

$$\tilde{\sigma}_i^2 = \frac{1}{n_i - 1} \sum_{j=1}^{n_i} \xi(\mathbf{A}_j^i)^2 - \frac{1}{n_i(n_i-1)} \times \Big(\sum_{j=1}^{n_i} \xi(\mathbf{A}_j^i) \Big)^2. \tag{16}$$

Consequently,

$$\tilde{W}_i^r = 0 \text{ if } i \in \tilde{K} \tag{17}$$

and

$$\tilde{W}_i^r = \frac{\rho_i \tilde{\sigma}_i}{\sum_{j \in \tilde{K}^c} \rho_j \tilde{\sigma}_j} \big(1 + \frac{\gamma_1(n)}{\gamma_2(n)} \sum_{j\in\tilde{K}^c} w_j\big) - \frac{\gamma_1(n)}{\gamma_2(n)} w_i \text{ if } i \in \tilde{K}^c \tag{18}$$

where $\tilde{K}$ and $\tilde{K}^c$ are estimated replacing the deterministic sequence $(\sigma_i)_{i=1,\dots,m}$ by the random sequence $(\tilde{\sigma}_i)_{i=1,\dots,m}$ in algorithm 1. Besides, $\epsilon(n) \in (0,1]$ in equation 15 enables in the estimation step to draw some directions in a stratum even if the estimated allocation returns zero, which can give a correction to the bias brought by a wrong estimation of an allocation. For a limited number of simulations, we should take $\epsilon(n)$ as small as possible: we propose $\epsilon(n) = 2/(n\rho_i) = 2^{p+1}/n$.
We denote for any positive integer k:

$$P_{i,k} = \mathbb{E}(\xi(\mathbf{A}^i)^k)$$

and

$$B\big(\hat{P}_{f,r}^{ADS-2}\big) = \mathbb{E}\big(\hat{P}_{f,r}^{ADS-2}\big) - P_f$$

the bias of the estimator. Then, the recycling estimator is consistent and has the following bias due to the dependence of the optimal allocation estimation of the recycling estimator with the $\gamma_1(n)n$ first draws. The following result is proved in section 7.5

Proposition 5 *The recycling estimator, $\hat{P}_{f,r}^{ADS-2}$, is consistent and biased. Moreover, if we suppose γ_1 and ϵ do not depend on n then the bias is given by:*

$$B\big(\hat{P}_{f,r}^{ADS-2}\big) = -\sum_{i\in K^c} \rho_i (1-\epsilon)^2 \frac{n\gamma_2}{(N_i^{r,opt})^2} \times \Big(P_{i,3} - P_{i,1}P_{i,2} - 2P_{i,1}^2 (P_{i,2} - P_{i,1}^2) \Big) \times h_i(\gamma_1, \gamma_2, \sigma_{K^c}, \mathbf{w}_{K^c}, \rho_{K^c}) + o(\frac{1}{n}),$$

with $\sigma_{K^c} = (\sigma_i)_{i\in K^c}$, $\mathbf{w}_{K^c} = (w_i)_{i\in K^c}$ and $\rho_{K^c} = (\rho_i)_{i\in K^c}$ for some functions h_i defined in the proof of proposition 5.

This bias is of the order of $\frac{1}{n}$ and depends on the first, second and third order moments of the $\xi\big(A^i\big)$, so that this bias cannot be efficiently estimated with a small number of simulations.
Moreover, if we suppose that $\gamma_1(n)=\gamma_1$, then the variance of the ADS-2 recycling estimator can be expressed as proposed in the following proposition which is demonstrated in section 7.6.

Proposition 6 *If $\gamma_1(n)=\gamma_1$ and $\epsilon(n)=\epsilon$ then the recycling estimator, $\hat{P}^{ADS-2}_{f,r}$, has the following variance:*

$$Var\big(\hat{P}^{ADS-2}_{f,r}\big)=\frac{1}{n}\sum_{i\in K}\frac{\rho_i^2\sigma_i^2}{(1-\epsilon)\gamma_1 w_i+\epsilon\rho_i}+\frac{1}{n}\sum_{i\in K^c}\frac{\rho_i^2\sigma_i^2(1-\epsilon)\big(\gamma_1 w_i+\gamma_2\mathbb{E}(\tilde{W}_i^r)\big)+\epsilon\rho_i}{\Big((1-\epsilon)\big(\gamma_1 w_i+\gamma_2 w_i^{opt_2}\big)+\epsilon\rho_i\Big)^2}+o(\frac{1}{n}).$$

The first term of proposition 6 is the part of variance generated by the quadrants in which no additional simulations are required, i.e. the quadrants where the learning step simulations were not a retrospectively necessary. So, this part of variance is not minimized. The second term of proposition 6 is generated by the quadrants in which additional simulations are required. Thus the estimated optimal allocation can be obtained. This term of variance tends to be near the minimal variance, which is exactly reached when the allocation is the optimal one and $\epsilon=0$. We can notice that in the case where we can take $\epsilon=0$, e.g. when a sufficiently large number n is available, the variance expression becomes

$$Var\big(\hat{P}^{ADS-2}_{f,r}\big)=\frac{1}{n}\sum_{i\in K}\frac{\rho_i^2\sigma_i^2}{w_i\gamma_1}+\frac{\gamma_1\sum_{i\in K^c}w_i+\gamma_2\sum_{i\in K^c}\mathbb{E}(\tilde{W}_i^r)}{\big(\gamma_2+\gamma_1\sum_{i\in K^c}w_i\big)^2}\times\Big(\sum_{i\in K^c}\rho_i\sigma_i\Big)^2. \tag{19}$$

Then, we can prove (see section 7.7) the asymptotic convergence of the estimator to a Gaussian distribution under two types of assumptions.

Proposition 7
Firstly, if we suppose that $\epsilon(n)\to 0$, $\gamma_1(n)\to 0$ and $\gamma_1(n)n\to+\infty$ as $n\to+\infty$, then:

$$\sqrt{n}(\hat{P}^{ADS-2}_{f,r}-P_f)\xrightarrow[n\to+\infty]{\mathcal{L}}\mathcal{N}(0,\sigma^2_{opt_1}).$$

Secondly, if we suppose that $\epsilon(n)\to 0$ and $\gamma_1(n)=\gamma_1$, then:

$$\sqrt{n}(\hat{P}^{ADS-2}_{f,r}-P_f)\xrightarrow[n\to+\infty]{\mathcal{L}}\mathcal{N}(0,\sigma^2_{opt_2}).$$

The assumption $\gamma_1(n)\to 0$ and $\gamma_1(n)n\to+\infty$ implies $\gamma_1(n)=o\big(\gamma_2(n)\big)$ and corresponds to a case where a large number of simulations is available, then we must keep the major part of them for the estimation step. The assumption $\gamma_1(n)=\gamma_1$ corresponds to our framework: indeed, as the number of calls to the failure function is limited, it is the most appropriate assumption.
Then, in the case $\gamma_1(n)=\gamma_1$, from proposition 7, the non-asymptotic confidence interval is obtained by considering the error of the estimation as Gaussian with variance the non-asymptotic variance of the estimator given by proposition 6. We estimate the variance of the estimator given in proposition 6 by:

$$\tilde{\sigma}_r^2=\frac{1}{n}\Bigg[\sum_{i\in\tilde{K}}\frac{\rho_i^2\tilde{\sigma}_{i,r}^2}{(1-\epsilon)\gamma_1 w_i+\epsilon\rho_i}+\sum_{i\in\tilde{K}^c}\frac{\rho_i^2\tilde{\sigma}_{i,r}^2}{(1-\epsilon)\Big(\frac{\rho_i\tilde{\sigma}_{i,r}(\gamma_2+\gamma_1\sum_{j\in\tilde{K}^c}w_j)}{\sum_{j\in\tilde{K}^c}\rho_j\tilde{\sigma}_{j,r}}\Big)+\epsilon\rho_i}\Bigg] \tag{20}$$

with

$$\tilde{\sigma}_{i,r}^2=\frac{1}{N_i^r-1}\sum_{j=1}^{N_i^r}\xi(\mathbf{A}_j^i)^2-\frac{1}{N_i^r(N_i^r-1)}\times\Big(\sum_{j=1}^{N_i^r}\xi(\mathbf{A}_j^i)\Big)^2 \tag{21}$$

and $(\mathbf{A}_j^i)_{j=1,\dots,N_i^r}$ a family of i.i.d. random vectors with the distribution $\mathcal{L}(\mathbf{A}|\mathbf{A}\in q_i)$.
In the case $\gamma_1(n)\to 0$ and $\gamma_1(n)n\to+\infty$, i.e. when a large number of simulations is available, we can use the same strategy or suppose the error of the estimation as Gaussian with variance a direct estimation of the optimal variance $\sigma^2_{opt_1}$. We

can notice that, referring to the result of proposition 4, under the realistic assumption $\gamma_1(n) = \gamma_1$, the optimal standard deviation we achieve, σ_{opt_2}, will be larger than σ_{opt_1}.

3.3 The non-recycling ADS-2 estimator: definition and properties

The non-recycling ADS-2 estimator does not use the $\gamma_1(n)n$ first draws for the expectation estimation. It only uses the second part of draws, $\gamma_2(n)n$:

$$\hat{P}_{f,nr}^{ADS-2} = \sum_{i=1}^{m} \rho_i \frac{1}{N_i^{nr}} \sum_{j=1}^{N_i^{nr}} \xi(\mathbf{A}^i_{n_i+j}) \quad (22)$$

with

$$N_i^{nr} := \lfloor (1-\epsilon(n)) n_i^{nr} + \epsilon(n)\rho_i n \rfloor \quad (23)$$

with $n_i^{nr} = \left\lfloor \gamma_2(n) n \tilde{W}_i^{nr} \right\rfloor$ and $(\tilde{W}_i^{nr})_{i=1,\dots,m}$ given by proposition 2 replacing the $(\sigma_i)_{i\in I}$ by their estimations defined by equation (16).
Thus, for all $i \in I$:

$$\tilde{W}_i^{nr} = \frac{\rho_i \tilde{\sigma}_i}{\sum_{j=1}^{m} \rho_j \tilde{\sigma}_j}. \quad (24)$$

Besides, $\epsilon(n) \in (0,1]$, as for the recycling estimator, is the "conservative" allocation. The non-recycling estimator is consistent, unbiased and we can get an expression of its variance as shown in the following proposition which is proved in section 7.8.

Proposition 8 *The non-recycling ADS-2 estimator, $\hat{P}_{f,nr}^{ADS-2}$, is consistent, unbiased and its variance is given by:*

$$Var\left(\hat{P}_{f,nr}^{ADS-2}\right) = \frac{1}{n} \sum_{i=1}^{m} \rho_i^2 \sigma_i^2 \times \mathbb{E}\left(\frac{1}{\left(1-\epsilon(n)\right)\gamma_2(n) \frac{\rho_i \tilde{\sigma}_i}{\sum_{j=1}^{m} \rho_j \tilde{\sigma}_j} + \epsilon(n)\rho_i} \right)$$

for all n.

We also get the following Central Limit Theorem proved in section 7.9.

Proposition 9 *Firstly, if we assume that $\epsilon(n) \to 0$, $\gamma_1(n) \to 0$ and $\gamma_1(n)n \to +\infty$, then:*

$$\sqrt{n}(\hat{P}_{f,nr}^{ADS-2} - P_f) \xrightarrow[n\to+\infty]{\mathcal{L}} \mathcal{N}(0, \sigma_{opt_1}^2).$$

Secondly, if we assume that $\epsilon(n) \to 0$ and $\gamma_1(n) = \gamma_1$, then:

$$\sqrt{n}(\hat{P}_{f,nr}^{ADS-2} - P_f) \xrightarrow[n\to+\infty]{\mathcal{L}} \mathcal{N}(0, \frac{\sigma_{opt_1}^2}{1-\gamma_1}).$$

We notice that, in the case where $\gamma_1(n) = \gamma_1$, we do not achieve the optimal standard deviation, $\sigma_{opt_1}^2$: we loose a factor $\frac{\gamma_1}{1-\gamma_1}$ of it. To get the confidence interval, we assume that the error is Gaussian and estimate the variance of the estimator, given by proposition 8 , by:

$$\tilde{\sigma}_{nr}^2 = \frac{1}{n} \sum_{i=1}^{m} \frac{\rho_i^2 \tilde{\sigma}_{i,nr}^2}{[1-\epsilon(n)]\gamma_2(n) \frac{\rho_i \tilde{\sigma}_{i,nr}}{\sum_{j=1}^{m} \rho_j \tilde{\sigma}_{j,nr}} + \epsilon(n)\rho_i} \quad (25)$$

with

$$\tilde{\sigma}_{i,nr}^2 = \frac{1}{N_i^{nr}-1} \sum_{j=1}^{N_i^{nr}} \xi(\mathbf{A}_j^i)^2 - \frac{1}{N_i^{nr}(N_i^{nr}-1)} \times \Big(\sum_{j=1}^{N_i^{nr}} \xi(\mathbf{A}_j^i) \Big)^2 \quad (26)$$

and $(\mathbf{A}_j^i)_{j=1,\dots,N_i^{nr}}$ a family of i.i.d. random vectors with the distribution $\mathcal{L}(\mathbf{A}|\mathbf{A} \in q_i)$. Then, in the case $\gamma_1(n) = \gamma_1$, from proposition 9 and the previous standard deviation estimation, the confidence interval is obtained by considering the error of the estimation as Gaussian and by estimating the variance of this error by $\tilde{\sigma}_{nr}^2/(1-\gamma_1)$. In the case $\gamma_1(n) \to 0$ and $\gamma_1(n)n \to +\infty$, i.e. when a large number of simulations is available, we can use the same strategy or suppose the error of the estimation as Gaussian with variance a direct estimation of the optimal variance $\sigma_{opt_1}^2$. Also, we can prove that the asymptotic behavior of the estimation of the optimal allocation is Gaussian too (see section 7.10). Let us denote $\sigma^2 = (\sigma_i^2)_{i=1,\dots,m}$ and define the following functions:

$\phi_i(\mathbf{x}) = \frac{\rho_i \sqrt{x_i}}{\sum_{j=1}^{m} \rho_j \sqrt{x_j}}, \Psi(x) = \big(\phi_1(\mathbf{x}), ..., \phi_m(\mathbf{x})\big)$, $\nabla \phi_i$ the gradient of ϕ_i and J_Ψ the Jacobian matrix of Ψ. We denote for $i \in I$:

$$\delta_i^2 = \frac{D_i}{\gamma_1 w_i}$$

with $D_i = \big(P_{i,4} - 4P_{i,1}P_{i,3} + 8P_{i,1}^2 P_{i,2} - P_{i,2}^2 - 4P_{i,1}^4\big)$. Finally, we denote by Σ the diagonal matrix with $(\delta_1^2, ..., \delta_m^2)$ its diagonal vector.

Proposition 10 *We assume that* $\gamma_1(n) = \gamma_1$ *and* $\epsilon(n) \to 0$ *then:*

$$\sqrt{n}(\tilde{W}_i^{nr} - w_i^{opt_1}) \xrightarrow[n \to +\infty]{\mathcal{L}} \mathcal{N}(0, \eta_i)$$

with $\eta_i = \nabla \phi_i(\sigma^2) \Sigma \nabla \phi_i^t(\sigma^2)$*, and*

$$\sqrt{n}(\tilde{\mathbf{W}}^{\mathbf{nr}} - \mathbf{w_{opt_1}}) \xrightarrow[n \to +\infty]{\mathcal{L}} \mathcal{N}(0, \Gamma)$$

with $\Gamma = J_\Psi(\sigma^2) \Sigma J_\Psi^t(\sigma^2)$.

In particular we have:

$$\Sigma_{ii} = \eta_i = \frac{D_1}{\gamma_1 w_1}\left[\frac{\rho_i (w_i^{opt_1})^2}{\sigma_1 \rho_i \sigma_i}\right]^2 + ... + \frac{D_{i-1}}{\gamma_1 w_{i-1}}\left[\frac{\rho_i (w_i^{opt_1})^2}{\sigma_{i-1} \rho_i \sigma_i}\right]^2 + \frac{D_i}{\gamma_1 w_i}\left[\frac{w_i^{opt_1}(1 - w_i^{opt_1})}{2\sigma_i^2}\right]^2 + \frac{D_{i+1}}{\gamma_1 w_{i+1}}\left[\frac{\rho_i (w_i^{opt_1})^2}{\sigma_{i+1} \rho_i \sigma_i}\right]^2 + ... + \frac{D_m}{\gamma_1 w_m}\left[\frac{\rho_i (w_i^{opt_1})^2}{\sigma_m \rho_i \sigma_i}\right]^2. \quad (27)$$

4 An improvement of the ADS-2 method

When the physical dimension grows, the number of strata of the ADS-2 method increases exponentially: indeed, in dimension p, the number of quadrants is 2^p. As a minimum of simulations is required to explore each quadrant, the number of directional simulations needed can be too large with respect to the restricted number of simulations we are limited to (C_0). Now, the idea is to get, with the simulations performed in the first step (learning step), a sort of the random variables in function of their influence on the failure event. Then, we only stratify the most important ones. We use the simulations done in the first step to detect the important variables and to estimate the optimal allocation in the new strata. In the second step, we get the final estimation with this allocation. To determine if a random variable will be stratified, we propose the following method. We first index the quadrants. The input index $k \in 1, ..., p$ is given the tag: i_k which takes its values in $\{-1, 1\}$ and corresponds to the input sign. Thus, each quadrant $i \in \{1, ..., m\}$ is characterized by a p-uple $\mathbf{i} = (i_1, ..., i_p)$ (see Figure 4.1). From now on, for the strata indexation, we will use, in an equivalent way, either the indexation $i \in \{1, ..., m\}$ or its associated multi-index $\mathbf{i} = (i_1, ..., i_p) \in \{-1, 1\}^p$. Then, we define the sequence $\tilde{\mathbf{T}} = (\tilde{T}_k)_{k=1,...,p}$

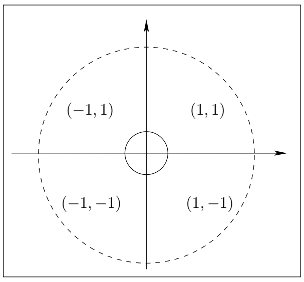

Figure 4.1 Two-dimensional strata indexation illustration.

by:

$$\tilde{T}_k = \sum_{i_l \in \{-1,1\}, l \neq k} \left| \tilde{P}_{(i_1,...,i_{k-1},-1,i_{k+1},...,i_p)} - \tilde{P}_{(i_1,...,i_{k-1},1,i_{k+1},...,i_p)} \right| \quad (28)$$

with $\tilde{P}_{(i_1,...,i_p)}$ the estimation of the expectation in the quadrant $(i_1, .., i_p)$ obtained during the learning step.

Thus, $\tilde{T}_k$ aggregates the differences of the expectations between the quadrants along dimension k. The larger $\tilde{T}_k$ is, the more influential the k-th input is. Then, we sort sequence $\tilde{\mathbf{T}}$ by decreasing order and we decide to stratify only over the $p' < p$ first dimensions. The other inputs being simulated without stratification.

Next, we need to estimate the optimal allocation to be achieved in the new $m' = 2^{p'}$ hyper-quadrants. Let us denote $(q_{p,i})_{i=1,...,m}$ the partition of $\mathbb{R}^p$ into quadrants and $(q_{p',i})_{i=1,...,m}$ the partition of $\mathbb{R}^{p'}$ into quadrants. We define $\tilde{\mu}$ the $p \times p'$ matrix such that:

$$\tilde{\mu}_{(i),i} = 1 \text{ for } i = 1, ..., p' \text{ else } 0. \quad (29)$$

Then, after the selection of the p' variables on which the stratification will be performed, the new strata are given by:

$$q_i(\tilde{\mu}) = \{x \in \mathbb{R}^p, \tilde{\mu}^t x \in q_{p',i}\} \quad \text{for } i = 1, ..., m'. \quad (30)$$

Now, as the new strata are defined, we need to estimate the optimal allocation to perform in these strata with the simulations already made. To do so, we use the same formulas as the ones presented in section 3.2, using once again the same set of simulations (of the first step). For $i \in \{1, ..., m'\}$, we have $\rho'_i = 1/m'$, the probability of being in the i-th quadrant $q_i(\tilde{\mu})$, which is independent of $\tilde{\mu}$ as the strata are quadrants. As previously seen, to an index $i \in \{1, ..., m'\}$, we can associate the p'-uple $\mathbf{i} = \big(i_{(1)}, ..., i_{(p')}\big) \in \{-1, 1\}^{p'}$. So, the number of directional simulations performed in the quadrant $q_i(\tilde{\mu})$ will be:

$$n_i(\tilde{\mu}) = \gamma_1(n)n \sum_{i_j \in \{-1,1\}, j \notin \{(1),...,(p')\}} w_{(i_1,...,i_p)},$$

hence for a proportional prior allocation $w_{(i_1,...,i_p)} = \rho_{(i_1,...,i_p)} = 1/m$ we get $n_i(\tilde{\mu}) = n'_i = \gamma_1(n)n\rho'_i$. Finally, for $i \in \{1, ..., m'\}$, we naturally define the estimated standard deviation, the estimated optimal allocation and the estimated optimal number of simulations in the quadrant $q_i(\tilde{\mu})$ by:

$$\tilde{\sigma}_i(\tilde{\mu})^2 = \frac{1}{n_i(\tilde{\mu}) - 1} \sum_{j=1}^{n_i(\tilde{\mu})} \xi(\mathbf{A}^i_j)^2 - \frac{1}{n_i(\tilde{\mu})(n_i(\tilde{\mu}) - 1)} \Big(\sum_{j=1}^{n_i(\tilde{\mu})} \xi(\mathbf{A}^i_j)\Big)^2 \quad (31)$$

with $(\mathbf{A}^i_j)_{j=1,...,n_i(\tilde{\mu})}$ a family of i.i.d random vectors with the distribution $\mathcal{L}\big(\mathbf{A}|\mathbf{A} \in q_i(\tilde{\mu})\big)$,

$$\tilde{W}^{nr}_i(\tilde{\mu}) = \frac{\rho'_i \tilde{\sigma}_i(\tilde{\mu})}{\sum_{j=1}^{m'} \rho'_j \tilde{\sigma}_j(\tilde{\mu})}, \quad (32)$$

and

$$N^{nr}_i(\tilde{\mu}) =: \Big\lfloor \big(1 - \epsilon(n)\big)\gamma_2(n)n\tilde{W}^{nr}_i(\tilde{\mu}) + \epsilon(n)\rho'_i n \Big\rfloor. \quad (33)$$

Consequently, the final ADS-2^+ estimator is given by:

$$\hat{P}^{ADS-2^+}_{f,nr}(\tilde{\mu}) = \sum_{i=1}^{m'} \rho'_i \frac{1}{N^{nr}_i(\tilde{\mu})} \sum_{j=1}^{N^{nr}_i(\tilde{\mu})} \xi(\mathbf{A}^i_j) \quad (34)$$

with $(\mathbf{A}^i_j)_{j=1,...,N^{nr}_i(\tilde{\mu})}$ a family of i.i.d random vectors with the distribution $\mathcal{L}\big(\mathbf{A}|\mathbf{A} \in q_i(\tilde{\mu})\big)$.

5 Numerical applications

5.1 Numerical illustration of the theoretical results on an academic example

In this section, we test numerically the efficiency of the ADS-2 method on an academic example. We determine the behavior of the ADS-2 method in the case where the failure surface is essentially concentrated in one quadrant but with some part running over other quadrants. So, we test our ADS-2 method considering the hyperplane failure surfaces:

$$H_2 : \frac{1}{4}x_1 + \sum_{i=2}^{p} x_i = k$$

with p the physical space dimension. To determine the constant k corresponding to a given failure probability, P_f, we just need to calculate some appropriate Gaussian quantiles.
Let us denote by b the radius of the sphere beyond which the probabilities are insignificant in comparison with the seeked failure probability. It is defined in the p-dimensional Gaussian space by:

$$b = \sqrt{\chi^{-1}_{2,p}(1 - mP_f)},$$

with $\chi^{-1}_{2,p}$ the inverse chi-square cdf with p degrees of freedom and m some margin, for example $m = 10^{-1}$. The failure probability to estimate is

$P_f = 10^{-6}$, the physical space dimension considered is $p = 3$, the percentage of directional draws allocated in the first step of the ADS-2 method, γ_1, takes its values between 15% and 85%:

$$\gamma_1 \in \{15\%, 50\%, 85\%\}$$

and the total number of directional draws, n, takes its values between 100 and 5000:

$$n \in \{100, 120, 140, 160, 180, 200, 240, 260, 280, 300, 500, 750, 1000, 1250, 1500, 2000, 3500, 5000\}.$$

Also, we denote by N the total number of runs of the method. $\hat{P}_{f,M,k}^{ADS-2}$ and $\hat{\sigma}_{M,k}^{ADS-2}$ are the k-th expectation estimation and the k-th standard deviation estimation (non-asymptotic) with the estimator $M = r$ or nr for respectively the recycling or the non-recycling estimator.
To begin with, we present in Figure 5.1 the mean relative bias of both recycling and non-recycling estimators defined by:

$$\hat{B} = \frac{1}{N} \sum_{k=1}^{N} \hat{B}_k$$

with

$$\hat{B}_k = \frac{\hat{P}_{f,M,k}^{ADS-2} - P_f}{P_f}.$$

We notice, as predicted respectively in propositions 5 and 8, that the recycling estimator is biased and the non-recycling estimator is unbiased. The bias of the recycling estimator is significant for small values of n. For large values of n the bias of the recycling estimator does not converge to zero. We choose an ϵ equal to zero, hence we highlight the constant bias which can be brought by neglecting this parameter.

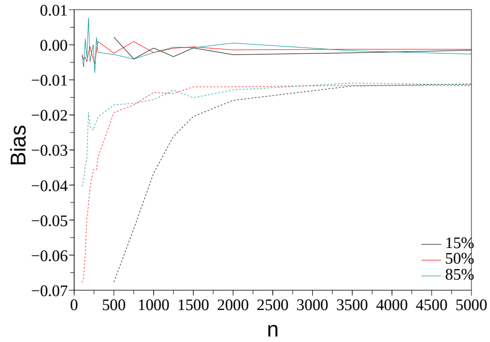

Figure 5.1 The mean relative bias of the recycling (dashed line) and non-recycling (full line) estimators depending on n and γ_1. $N = 1000$.

Then, we present in Figure 5.2 the histogram of both recycling and non-recycling estimations for a large value of n. Hence, under the hypothesis $\gamma_1(n) = \gamma_1$, we compare the asymptotic results given by the propositions 7 and 9. In other words, Figure 5.2 gives an idea of the asymptotic error of estimations, which are respectively $\sigma_{opt_2}^2$ and $\sigma_{opt_1}^2/(1-\gamma_1)$ for the recycling and the non-recycling estimators. The recycling error seems to be a little smaller than the non-recycling one.

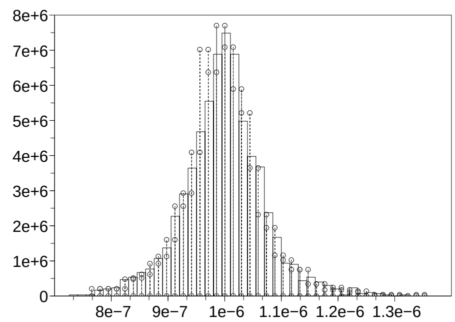

Figure 5.2 Histograms of the estimations of the failure probability performed with the non-recycling (full line) and recycling (dash bar with circle) estimators. $N = 1000$. $\gamma_1 = 50\%$. $n = 2000$.

Figures 5.3 and 5.4 respectively present the behavior of the non-asymptotic standard deviation of the recycling and the non-recycling estimators, respectively given by equations 20 and 25, depending on n. It is compared with a direct estimation of the optimal asymptotic standard deviation which is given for the recycling estimator by proposition 3 and estimated by

$$\frac{1}{n}\left[\sum_{i\in\tilde{K}}\frac{\rho_i^2\tilde{\sigma}_{i,r}^2}{\gamma_1(n)w_i}+\frac{1}{\gamma_2(n)+\gamma_1(n)\sum_{i\in\tilde{K}^c}w_j}\times\left(\sum_{i\in\tilde{K}^c}\rho_i\tilde{\sigma}_{i,r}\right)^2\right]$$

and for the non-recycling estimator by proposition 2 and estimated by

$$\frac{1}{n\gamma_2}\left(\sum_{i=1}^{m}\rho_i\tilde{\sigma}_{i,nr}\right)^2.$$

Also, Figures 5.3 and 5.4 provide the empirical estimator of the standard deviation defined by

$$\sqrt{\frac{1}{N}\sum_{k=1}^{N}(\hat{P}_{f,M,k}^{ADS-2})^2-(\frac{1}{N}\sum_{k=1}^{N}\hat{P}_{f,M,k}^{ADS-2})^2}.$$

with $M = r$ or nr according to the method considered. In both cases, we observe the convergence of the three previous indicators to the same limit when n tends to get large. Moreover, we notice that the empirical estimation begins to get close to the asymptotic and non-asymptotic estimations for relatively small values of n which justify the use of these ones in the case of a limited value for n.

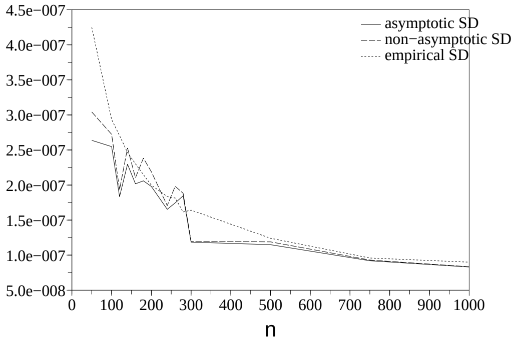

Figure 5.3 Asymptotic, non-asymptotic and empirical estimations of the Standard Deviation (SD) of the recycling estimator depending on n. $\gamma_1 = 50\%$. $N = 1000$.

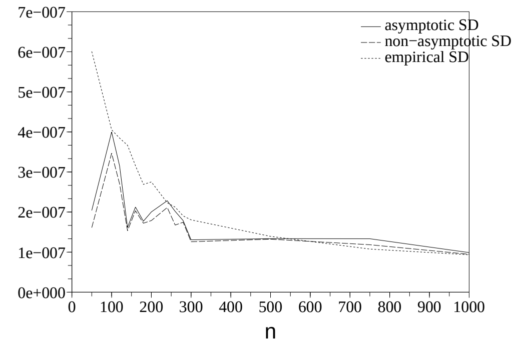

Figure 5.4 Asymptotic, non-asymptotic and empirical estimation of the Standard Deviation (SD) of the non-recycling estimator depending on n. $\gamma_1 = 50\%$. $N = 1000$.

Figure 5.5 presents the histogram of the non-recycling estimations, for $N = 1000$ runs of the method, under two hypotheses: on the one hand $\gamma_1(n) \to 0$ and $\gamma_1(n)n \to +\infty$ and on the other hand $\gamma_1(n) = \gamma_1$. As n is taken large, Figure 5.5 illustrates proposition 9. Indeed we notice, as predicted, a larger error of estimation under the hypothesis $\gamma_1(n)$ independent of n and equal to γ_1. Figure 5.6 presents the same analysis for the recycling estimator. It illustrates proposition 7 and, as predicted, a larger error of estimation is reached under the hypothesis $\gamma_1(n) = \gamma_1$.

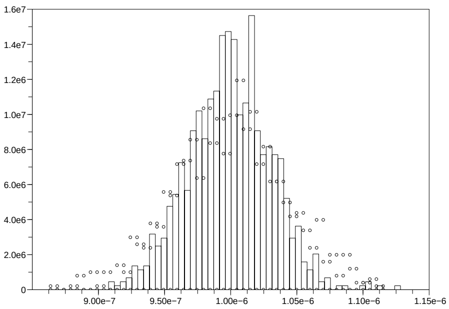

Figure 5.5 Histogram of the non-recycling estimator for $\gamma_1 = 7\%$ (full line) and $\gamma_1 = 50\%$ (circle). $N = 1000$.

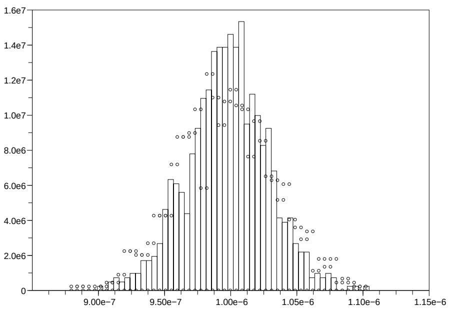

Figure 5.6 Histogram of the recycling estimator for $\gamma_1 = 7\%$ (full line) and $\gamma_1 = 50\%$ (circle). $N = 1000$.

5.2 Comparison of ADS with the directional simulation and the subset simulation methods

In this section, we apply the ADS methods to the hyperplanes H_1 and H_3 defined by

$$H_1 : \sum_{i=1}^{p} x_i = k_1 \text{ and } H_3 : x_p = k_3.$$

H_3 has one influential variable and H_1 has all its variables influential. We define the estimated percentage of estimations fallen in the estimated two-sided symmetric 95% confidence interval as:

$$\hat{PCI} = \frac{100}{N} \sum_{k=1}^{N} \mathbb{1}_{P_f \in [\hat{P}^-_{M,k};\hat{P}^+_{M,k}]},$$

with

$$\hat{P}^{\pm}_{M,k} = \hat{I}^{ADS-2}_{M,k} \pm \alpha_{97,5\%} \frac{\hat{\sigma}^{ADS-2}_{M,k}}{\sqrt{n}},$$

where $\alpha_{97,5\%}$ is the $97,5\%$ Gaussian quantile. We compare in tables 5.1 and 5.2 the ADS-2$^+$ method with the Subset Simulation method (SS) which is built to overcome the curse of dimensionality and responds to the four constraints presented in section 1 (see [1]). Also, the results obtained with the ADS-2 method are presented. The SS results have been obtained using the open-source Matlab toolbox: FERUM (Finite Element Reliability Using Matlab) version 4.0.

Method	n	NG	$\hat{CV}$ (%)	$\hat{PCI}$
SS	500	3300	60	77
SS	700	4600	51	88
SS	1000	6700	42	81
DS	600	3100	42	86
DS	900	4501	35	90
ADS-2	1200	3219	55	61
ADS-2	1750	4675	46	61
ADS-2	2550	6760	39	76
ADS-2$^+$ ($p' = 3$)	1200	3139	41	78
ADS-2$^+$ ($p' = 3$)	1800	4799	31	83
ADS-2$^+$ ($p' = 2$)	1500	3962	30	90
ADS-2$^+$ ($p' = 1$)	1500	3957	28	92

Table 5.1 Results with hyperplane H_3. $P_f = 10^{-6}$. $p = 5$. $N = 100$.

Method	n	NG	$\hat{CV}$ (%)	$\hat{PCI}$
SS	300	1989	65	86
SS	500	3330	51	88
SS	700	4630	43	85
SS	1000	6800	36	93
DS	200	3154	62	70
DS	400	6672	50	82
DS	600	10100	42	82
ADS-2	300	2033	16	90
ADS-2	500	3340	13	93
ADS-2	650	4426	11	96
ADS-2	750	5885	10	92
ADS-2$^+$ ($p' = 3$)	700	2487	21	92
ADS-2$^+$ ($p' = 3$)	1200	4333	16	94

Table 5.2 Results with hyperplane H_1. $P_f = 10^{-6}$. $p = 5$. $N = 100$.

Table 5.1 confirms the efficiency of the ADS-2$^+$ method in comparison with the ADS-2 method. Moreover, in comparison with the SS method and for approximatively the same $\hat{CV}$ and $\hat{PCI}$, the

ADS-2^+ method ($p' = 3$) enables to reduce the number of calls to the failure function by approximatively a factor 2. Also, for a better choice of p' (equal to 1), we can reduce the $\hat{CV}$ of 25% for approximatively the same number of calls to the failure function. As previously demonstrated in section 2.3, we emphasis the fact that hyperplan H_3 corresponds to the worst case for the ADS-2 and ADS-2^+ ($p' = 3$) methods. Finally, the same study has been performed on H_1 and H_3 for p going from 5 to 8 and $P_f = 10^{-6}$ and 10^{-8}. As predictable, the SS method is completely robust with respect to the increase of the dimension and gives almost exactly the same results. On H_1, the ADS-2 method always outperforms the SS method: for instance in dimension 7 for a failure probability of 10^{-8}, we divide by a factor 2 the $\hat{CV}$ with the half number of calls to the failure function used for SS (8000) while keeping a good $\hat{PCI}$. In the worst case, i.e. when considering H_3, $P_f = 10^{-8}$ and $p = 7$, the results between SS and ADS-2^+ are equivalent. Finally, we can see in tables 5.1 and 5.2 that the ADS methods outperform the simple directional simulation method, slightly on H_3 and completely on H_1.

6 Discussion and conclusion

The ADS-2 strategy concentrates the runs into the most important parts of the sample space, which results in an asymptotic optimal error variance. For a limited number of simulations, the theoretical and numerical results we have presented show that a relevant solution to get a confidence interval on the estimation is to consider the error Gaussian and to consider the non-asymptotic estimation of the variance of the estimator.

Furthermore, when the number of simulations is limited, the most natural assumption will be to choose $\gamma_1(n) = \gamma_1$, in order to have a sufficient number of draws in both learning and estimation steps of the ADS-2 method. A numerical study over this parameter has been performed in Munoz Zuniga et al. [28]. Another point is to make a choice between the recycling and the non-recycling estimators. Intuitively, the first choice would be for the recycling one, since re-using the first step draws seems to be more efficient to estimate the failure probability. However, the problem is that this recycling estimator is biased and this bias is not easy to estimate and correct with a limited number of simulations, which is conflicting with the robustness we are looking for. Consequently, the unbiased non-recycling estimator seems to be a better choice: even if we loose some factor in the standard deviation, we know exactly how much. Also in order to avoid a constant bias on the estimation, it is important to choose a non zero value for the parameter $\epsilon(n)$ which enables, during the estimation step, to perform some simulations in the quadrants which are not detected as important ones in the learning step. Finally, we compare the ADS-2 and ADS-2^+ methods to the subset simulation method which is one of the most relevant method to use in the context described in section 1. The results show that in the considered configurations the ADS-2^+ method outperforms the subset simulation method. Hence, the ADS methods are very efficient when the following conditions are met: a number of calls to the failure function of a few thousand, an order of magnitude of the failure probability less than 10^{-4}, a dimension less than 7, no regularity assumption on the failure function, and a need for explicit and trustworthy error controls. To summarize, the methods we have proposed respond substantially to the initial constraints for a realistic space dimension often met in mechanical issues for instance. To our opinion, it may not be unreasonable to concentrate onto at most half a dozen important variables when looking for improved accuracy in the prediction of rare failure probabilities. For larger sets of uncertain variables that may be involved in upstream stages of risk analysis, undertaking prior physical or numerical sensitivity studies may prove more relevant; remember also that the accuracy in the computation of failure probability makes sense only if there is high confidence in the description of the input uncertainty distributions.

7 Proofs

7.1 Proof of proposition 1

The consistency of the estimator is obtained by a direct application of the the Law of Large Numbers (LLN) as ξ is bounded. The total probability formula implies that $P_f = \sum_{i=1}^{m} \rho_i P_i$, so that:

$$\sqrt{n}\big(\hat{P}_f^{DSS} - P_f\big) = \sqrt{n}\big(\sum_{i=1}^{m} \rho_i(\hat{P}_i - P_i)\big)$$

$$= \sum_{i=1}^{m} \frac{\sqrt{n}}{\sqrt{n_i}} \rho_i \sqrt{n_i}(\hat{P}_i - P_i)$$

$$= \sum_{i=1}^{m} \frac{\rho_i}{\sqrt{w_i}} \sqrt{n_i}(\hat{P}_i - P_i).$$

Indeed, as $n_i = nw_i$, we have $\frac{\sqrt{n}}{\sqrt{n_i}} = \frac{1}{\sqrt{w_i}}$ (neglecting the rounding approximation error). Now, let us consider the column vector

$$\rho := \big(\frac{\rho_i}{\sqrt{w_i}}\big)_{i\in I}.$$

Using the characteristic function, we prove that:

$$\begin{pmatrix} \sqrt{n_1}(\hat{P}_1 - P_1) \\ \vdots \\ \sqrt{n_m}(\hat{P}_m - P_m) \end{pmatrix} \xrightarrow[n\to+\infty]{\mathcal{L}} \mathcal{N}(0, \Sigma)$$

where Σ is a diagonal $m\times m$ matrix with $(\sigma_1^2, ..., \sigma_m^2)$ the main diagonal.
Then, the Slutsky theorem implies that:

$$\rho \begin{pmatrix} \sqrt{n_1}(\hat{P}_1 - P_1) \\ \vdots \\ \sqrt{n_m}(\hat{P}_m - P_m) \end{pmatrix} \xrightarrow[n\to+\infty]{\mathcal{L}} \mathcal{N}(0, \rho^t \Sigma \rho)$$

with $\rho^t \Sigma \rho = \sum_{i=1}^{m} \frac{\rho_i^2 \sigma_i^2}{w_i}$, which completes the proof. □

7.2 Proof of proposition 2

Let $L(\mathbf{w}, \lambda)$ be the Lagrangian function associated with the problem (E_1) with $\mathbf{w} = (w_i)_{i\in I}$. Let us denote $c_i := \frac{\rho_i^2 \sigma_i^2}{n}$, then, for all $i \in I$

$$L(\mathbf{w}, \lambda) = \sum_{i=1}^{m} \frac{c_i}{w_i} + \lambda\big(\sum_{i=1}^{m} w_i - 1\big)$$

by differentiating we find:

$$\frac{\partial L}{\partial w_i}(\mathbf{w}, \lambda) = -\frac{c_i}{w_i^2} + \lambda = 0 \Leftrightarrow w_i = \frac{\sqrt{c_i}}{\sqrt{\lambda}} \tag{35}$$

and

$$\frac{\partial L}{\partial \lambda}(\mathbf{w}, \lambda) = 0 \Leftrightarrow \sum_{i=1}^{m} w_i = 1. \tag{36}$$

Summing the w_i in equation (35) and using (36), we have:

$$\frac{1}{\lambda} = \frac{1}{\sum_{i=1}^{m} \sqrt{c_i}}.$$

Substituting this into (35), we get for all $i \in I$:

$$w_i = \frac{\sqrt{c_i}}{\sum_{i=1}^{m} \sqrt{c_i}} \geq 0$$

and the constraint is satisfied, which completes the proof. □

7.3 Proof of proposition 3

We use the Lagrangian method. The Lagrangian function for $(\mathbf{w}, \lambda) \in \mathbb{R}_+^m \times \mathbb{R}$ is:

$$L(\mathbf{w}, \lambda) = \sum_{i=1}^{m} \frac{c_i}{\gamma_1(n)\beta_i + \gamma_2(n) w_i} + \lambda\big(\sum_{i=1}^{m} w_i - 1\big)$$

$$= \sum_{i=1}^{m} \big[\frac{c_i}{\gamma_1(n)\beta_i + \gamma_2(n) w_i} + \lambda w_i\big] - \lambda$$

with $c_i = \frac{\rho_i^2 \sigma_i^2}{n}$.
We choose to solve the dual problem: we look for the couple $(\mathbf{w}(\lambda^*), \lambda^*)$ such that:

$$L(\mathbf{w}(\lambda^*), \lambda^*) = \max_{\lambda\geq 0} \min_{\mathbf{w}\in\mathbb{R}_+^m} L(\mathbf{w}, \lambda)$$

so that $\mathbf{w}(\lambda^*)$ will be the solution of the initial constrained problem.
First, for a fixed λ, we want to minimize L. Let

$$\mathbf{w}(\lambda) = \underset{\mathbf{w}\in\mathbb{R}_+^m}{argmin} L(w, \lambda)$$

be the solution.
Minimizing L is equivalent to minimize, for all $i \in I$, the function $h_i(.,\lambda) : x \to \frac{c_i}{\gamma_1(n)\beta_i+\gamma_2(n)x} + \lambda x$ for $x \in \mathbb{R}_+$.
That is why we need to study the behavior of $h_i(.,\lambda)$. The derivative of $h_i(.,\lambda)$ has the same sign as: $\lambda(\gamma_1(n)\beta_i + \gamma_2(n)x)^2 - c_i\gamma_2(n)$, and three cases are possible:

(a) $\lambda \leq 0$: then, for all i, $\lambda \to h_i(.,\lambda)$ is decreasing and $w_i(\lambda) = +\infty$.

(b) $0 < \lambda \leq \frac{c_i\gamma_2(n)}{\beta_i^2\gamma_1(n)^2}$: then, $h_i(.,\lambda)$ reaches its unique minimum at $w_i(\lambda) = \frac{\sqrt{c_i}}{\sqrt{\gamma_2(n)\lambda}} - \frac{\gamma_1(n)\beta_i}{\gamma_2(n)}$.

(c) $\lambda > \frac{c_i\gamma_2(n)}{\beta_i^2\gamma_1(n)^2}$: then, $h_i(.,\lambda)$ is increasing and $w_i(\lambda) = 0$.
Now, we must look for the $\lambda \in (0,+\infty)$ which maximizes $L(w(\lambda),\lambda)$. Analyzing the three previous cases, we see that this maximum will be obtained when $\lambda > 0$, so that for $\lambda \in (0,+\infty)$ we want to maximize:

$$L(\mathbf{w}(\lambda),\lambda) = \sum_{i=1}^{m} \big[1_{\lambda > \frac{c_i\gamma_2(n)}{\beta_i^2\gamma_1(n)^2}} \frac{c_i}{\gamma_1(n)\beta_i} + 1_{\lambda \leq \frac{c_i\gamma_2(n)}{\beta_i^2\gamma_1(n)^2}} (2\frac{\sqrt{c_i\lambda}}{\sqrt{\gamma_2(n)}} - \lambda\frac{\gamma_1(n)\beta_i}{\gamma_2(n)})\big] - \lambda.$$

Then:

$$\partial_\lambda L(\mathbf{w}(\lambda),\lambda) = \sum_{i=1}^{m} \big[1_{\lambda \leq \frac{c_i\gamma_2(n)}{\beta_i^2\gamma_1(n)^2}} (\frac{\sqrt{c_i}}{\sqrt{\lambda\gamma_2(n)}} - \frac{\gamma_1(n)\beta_i}{\gamma_2(n)})\big] - 1.$$

The function $\lambda \to \partial_\lambda L(\mathbf{w}(\lambda),\lambda)$ is continuous and equal to -1 for $\lambda > \max_{i\in\{1,\ldots,m\}} \frac{c_i\gamma_2(n)}{\beta_i^2\gamma_1(n)^2}$ and strictly decreasing on $(0, \max_{i\in\{1,\ldots,m\}} \frac{c_i\gamma_2(n)}{\beta_i^2\gamma_1(n)^2}]$. We also have:

$$\partial_\lambda L(\mathbf{w}(\lambda),\lambda) \xrightarrow{\lambda\to 0} +\infty.$$

Consequently, $\partial_\lambda L(\mathbf{w}(\lambda),\lambda) = 0$ for a unique $\lambda^* \in (0, \max_{i\in\{1,\ldots,m\}} \frac{c_i\gamma_2(n)}{\beta_i^2\gamma_1(n)^2})$, and the function $L(\mathbf{w}(\lambda),\lambda)$ of the variable λ, reaches its unique maximum at λ^*. Let us search λ^*.
For $i \in I$, we sort by increasing order the quantities: $\frac{c_i\gamma_2(n)}{\beta_i^2\gamma_1(n)^2}$ and index the ordered sequence with (i). This ordered sequence gives a partition of the initial interval research and we now look for the sub-interval containing λ^*. Therefore, if

$$\partial_\lambda L(\mathbf{w}(\frac{c_{(i)}\gamma_2(n)}{\beta_{(i)}^2\gamma_1(n)^2}), \frac{c_{(i)}\gamma_2(n)}{\beta_{(i)}^2\gamma_1(n)^2}) < 0$$

for all i, we set $i^* = 0$ (this step can be done before the ordering step), otherwise we set i^* as the index such that:

$$\partial_\lambda L(\mathbf{w}(\frac{c_{i^*}\gamma_2(n)}{\beta_{i^*}^2\gamma_1(n)^2}), \frac{c_{i^*}\gamma_2(n)}{\beta_{i^*}^2\gamma_1(n)^2}) \geq 0$$

and

$$\partial_\lambda L(\mathbf{w}(\frac{c_{i^*+1}\gamma_2(n)}{\beta_{i^*+1}^2\gamma_1(n)^2}), \frac{c_{i^*+1}\gamma_2(n)}{\beta_{i^*+1}^2\gamma_1(n)^2}) < 0.$$

So, if $i^* = 0$ then

$$\lambda^* \in (0, \frac{c_{(1)}\gamma_2(n)}{\beta_{(1)}^2\gamma_1(n)^2})$$

otherwise

$$\lambda^* \in [\frac{c_{i^*}\gamma_2(n)}{\beta_{i^*}^2\gamma_1(n)^2}, \frac{c_{i^*+1}\gamma_2(n)}{\beta_{i^*+1}^2\gamma_1(n)^2}).$$

On this interval, we have:

$$\partial_\lambda L(\mathbf{w}(\lambda),\lambda) = \sum_{i=i^*+1}^{m} \big[(\frac{\sqrt{c_{(i)}}}{\sqrt{\lambda\gamma_2(n)}} - \frac{\gamma_1(n)\beta_{(i)}}{\gamma_2(n)})\big] - 1.$$

Moreover, we know that $\partial_\lambda L(\mathbf{w}(\lambda^*),\lambda^*) = 0$, which is equivalent to:

$$\frac{1}{\sqrt{\lambda^*\gamma_2(n)}} = \frac{1 + \frac{\gamma_1(n)}{\gamma_2(n)}\sum_{j=i^*+1}^{m}\beta_{(j)}}{\sum_{j=i^*+1}^{m}\sqrt{c_{(j)}}}. \quad (37)$$

Now, if $i^* = 0$, then $\lambda^* < \frac{c_{(i)}\gamma_2(n)}{\beta_{(i)}^2\gamma_1(n)^2}$ for all $i \in I$ and if $i^* \neq 0$, then $\lambda^* \geq \frac{c_{(i)}\gamma_2(n)}{\beta_{(i)}^2\gamma_1(n)^2}$ is equivalent to $i \leq i^*$. Thus, according to these equivalences, re-injecting (37) in the cases (b) and (c), we get:

$$w_{(i)}(\lambda^*) = 0 \quad \text{if} \quad i \leq i^*$$

and

$$w_{(i)}(\lambda^*) = \frac{\sqrt{c_{(i)}}}{\sum_{j=i^*+1}^{m}\sqrt{c_{(j)}}}\Big(1 + \frac{\gamma_1(n)}{\gamma_2(n)}\sum_{j=i^*+1}^{m}\beta_{(j)}\Big) - \frac{\gamma_1(n)}{\gamma_2(n)}\beta_{(i)} \quad \text{if} \quad i > i^*$$

and clearly $\sum_{i=1}^{m} w_{(i)}(\lambda^*) = 1$, so that the constraint is satisfied.
Finally, we have:

$$L(\mathbf{w}(\lambda^*), \lambda^*) = \max_{\lambda \in \mathbb{R}} \min_{\mathbf{w} \in [0,1]^m} L(\mathbf{w}, \lambda)$$

and $\mathbf{w}(\lambda^*)$ is the unique solution of (E_2).
As the initial function we want to minimize is a convex differentiable function and as the constraints are linear, the saddle point found verifies the Karush-Kuhn-Tucker (KKT) conditions and consequently is the unique solution of our problem. We refer to Pedregal (2004) [31] for the KKT conditions.
We can also look for a criterion to determine i^*. Using the fact that the function $\lambda \to L(\mathbf{w}(\lambda), \lambda)$ is concave and the definition of i^*, we have the following equivalences:

$$i \leq i^* \Leftrightarrow \partial L(w(\frac{c_{(i)}\gamma_2(n)}{\beta_{(i)}^2 \gamma_1(n)^2}), \frac{c_{(i)}\gamma_2(n)}{\beta_{(i)}^2 \gamma_1(n)^2}) \geq 0$$

$$\Leftrightarrow \frac{\beta_{(i)}}{\sqrt{c_{(i)}}} \geq \frac{\frac{\gamma_2(n)}{\gamma_1(n)} + \sum_{j=i+1}^{m} \beta_{(j)}}{\sum_{j=i+1}^{m} \sqrt{c_{(j)}}}$$

and

$$i^* = 0 \Leftrightarrow \frac{\beta_{(i)}}{\sqrt{c_{(i)}}} < \frac{\frac{\gamma_2(n)}{\gamma_1(n)} + \sum_{j=i+1}^{m} \beta_{(j)}}{\sum_{j=i+1}^{m} \sqrt{c_{(j)}}}, \quad \forall i \in I,$$

which completes the proof. □

7.4 Proof of proposition 4

The problem (E_2) can be written as follows:

$$(E_2) \begin{cases} \min_{w \in \mathbb{R}_+^m} \frac{1}{n} \left(\sum_{i=1}^{m} \frac{\rho_i^2 \sigma_i^2}{w_i} \right) \\ \text{under the constraints:} \\ \sum_{i=1}^{m} w_i = 1 \\ w_i \geq \gamma_1(n)\beta_i \end{cases}.$$

If we denote by

$$C_1 := \{w \in \mathbb{R}_+^m, \sum_{i=1}^{m} w_i = 1\}$$

the constraint of problem (E_1) and

$$C_2 := \{w \in \mathbb{R}_+^m, \sum_{i=1}^{m} w_i = 1 \text{ and } w_i \geq \gamma_1(n)\beta_i\}$$

the constraint of problem (E_2), then it is clear that $C_2 \subset C_1$ and consequently, for any objective function f, we have:

$$\min_{x \in C_1} f(x) \leq \min_{x \in C_2} f(x).$$

Thus,

$$\frac{\sigma_{opt_1}^2}{n} = \frac{1}{n} \big(\sum_{i=1}^{m} \rho_i \sigma_i \big)^2$$

$$\leq \frac{1}{n} \left(\sum_{i \in K} \frac{\rho_i^2 \sigma_i^2}{\gamma_1(n)\beta_i} + \frac{\left(\sum_{i \in K^c} \rho_i \sigma_i \right)^2}{\gamma_2(n) + \gamma_1(n) \sum_{i \in K^c} \beta_j} \right)$$

$$= \frac{\sigma_{opt_2}^2}{n}$$

which completes the proof. □

7.5 Proof of proposition 5

We recall that $n_i = \gamma_1(n) n w_i$, $n_i^r = \gamma_2(n) n \tilde{W}_i^r$ and $n_i^{nr} = \gamma_2(n) n \tilde{W}_i^{nr}$. Also,

$$N_i^r = (1 - \epsilon)\big(n_i + n_i^r\big) + \epsilon \rho_i n$$

and

$$N_i^{nr} = (1 - \epsilon) n_i^{nr} + \epsilon \rho_i n.$$

Now, we will denote by

$$N_i^{r,opt_2} = (1 - \epsilon)\big(n_i + n_i^{opt_2}\big) + \epsilon \rho_i n$$

and

$$N_i^{nr,opt_1} = (1 - \epsilon) n_i^{opt_1} + \epsilon \rho_i n$$

with

$$n_i^{opt_2} = \gamma_2(n) n w_i^{opt_2} \text{ and } n_i^{opt_1} = \gamma_2(n) n w_i^{opt_1}.$$

From now on, we denote the sigma-algebra generated by "the first step simulations" by:

$$\mathcal{F}_r^{\gamma_1} = \sigma\big(\mathbf{A}_1^1, ..., \mathbf{A}^1_{(1-\epsilon(n))n_1}, ..., \mathbf{A}_1^m, ..., \mathbf{A}^m_{(1-\epsilon(n))n_m}\big),$$

for the recycling strategy, and

$$\mathcal{F}_{nr}^{\gamma_1} = \sigma\big(\mathbf{A}_1^1, ..., \mathbf{A}^1_{n_1}, ..., \mathbf{A}_1^m, ..., \mathbf{A}^m_{n_m}\big),$$

for the non-recycling strategy. We also denote for any $k \in \mathbb{N}^*$:

$$P_{i,k} = \mathbb{E}(\xi(\mathbf{A}^i)^k)$$

and

$$\hat{P}_{i,k} = \frac{1}{(1-\epsilon)n_i} \sum_{j=1}^{(1-\epsilon)n_i} \xi(\mathbf{A}_j^i)^k$$

with $\mathbf{A}^i \sim \mathcal{L}(\mathbf{A}|\mathbf{A} \in q_i)$ and $\mathbf{A} \sim U(S_{p-1})$.
Let us begin with the consistency of the estimator. The recycling estimator can be expressed as

$$\hat{P}^{ADS-2}_{f,r} = \sum_{i=1}^m \rho_i \frac{N_i^{r,opt_2}}{N_i^r} \times \frac{1}{N_i^{r,opt_2}} \sum_{j=1}^{N_i^{r,opt_2}} \xi(\mathbf{A}_j^i) + R$$

with

$$R = \sum_{i=1}^m \frac{\rho_i}{N_i^r} \Bigg(\mathbb{1}_{\tilde{W}_i^r > w_i^{opt_2}} \sum_{j=N_i^{r,opt_2}+1}^{N_i^r} \xi(\mathbf{A}_j^i) - \mathbb{1}_{\tilde{W}_i^r \leq w_i^{opt_2}} \sum_{j=N_i^r+1}^{N_i^{r,opt_2}} \xi(\mathbf{A}_j^i) \Bigg).$$

Now, as ξ is bounded, we have for some constant C:

$$|R| \leq C \sum_{i=1}^m \frac{|N_i^r - N_i^{r,opt_2}|}{N_i^r}.$$

From the strong LLN, we can get that

$$\tilde{W}_i^r \xrightarrow[n\to+\infty]{a.s} w_i^{opt_2},$$

hence

$$N_i^r \xrightarrow[n\to+\infty]{a.s} N_i^{r,opt_2}$$

and

$$\frac{1}{N_i^{r,opt_2}} \sum_{j=1}^{N_i^{r,opt_2}} \xi(\mathbf{A}_j^i) \xrightarrow[n\to+\infty]{a.s} P_{i,1}.$$

Thus,

$$|R| \xrightarrow[n\to+\infty]{a.s} 0 \quad \text{and} \quad \hat{P}^{ADS-2}_{f,r} \xrightarrow[n\to+\infty]{a.s} I.$$

Now, let us determine the bias. We assume that $\gamma_1(n) = \gamma_1$ and $\epsilon(n) = \epsilon$. Let us consider the function g defined by $g(x) = \frac{1}{a+bx}$ with a, b and $x > 0$. Then, for some $x_0 > 0$, this function can be expressed as

$$g(x) = g(x_0) + g'(x_0)(x - x_0) + \frac{g^{(2)}(c)}{2}(x - x_0)^2$$

with c between x and x_0. Applying this Taylor-Lagrange expansion for $x = \tilde{W}_i^r$, $x_0 = w_i^{opt_2}$, $a = (1-\epsilon)\gamma_1 w_i + \epsilon\rho_i$ and $b = (1-\epsilon)\gamma_2$ then we get

$$\frac{n}{N_i^r} = \frac{n}{N_i^{r,opt_2}} - \frac{n^2(1-\epsilon)\gamma_2(\tilde{W}_i^r - w_i^{opt_2})}{(N_i^{r,opt_2})^2} + \frac{g^{(2)}(c_1)}{2}(\tilde{W}_i^r - w_i^{opt_2})^2. \tag{38}$$

Now, let us define, for $i \in K^c$, the functions $\Psi_i : \mathbb{R}_+^m \to \mathbb{R}$ by $\Psi_i(\mathbf{x}) = \frac{a_i x_i}{\sum_{j\in K^c} \rho_j x_j} - c_i$ then we have for some $\mathbf{x}^0 \in \mathbb{R}^m$:

$$\Psi_i(\mathbf{x}) = \Psi_i(\mathbf{x}^0) + \sum_{k=1}^m \frac{\partial \Psi_i}{\partial x_k}(\mathbf{x}^0)(x_k - x_k^0) + \frac{1}{2}\sum_{k,j=1}^m \frac{\partial^2 \Psi_i}{\partial x_k \partial x_j}(\mathbf{c})(x_k - x_k^0)(x_j - x_j^0)$$

with $\mathbf{c}$ on the segment between $\mathbf{x}$ and $\mathbf{x}^0$. We apply this result for all $i \in K^c$ with

$$a_i = \rho_i\big(1 + \frac{\gamma_1}{\gamma_2}\sum_{j\in K^c} w_j\big), \qquad c_i = \frac{\gamma_1}{\gamma_2} w_i,$$

$$\mathbf{x} = (\tilde{\sigma}_1, ..., \tilde{\sigma}_m) \quad \text{and} \quad \mathbf{x}^0 = (\sigma_1, ..., \sigma_m)$$

and under assumption (H), we get

$$\tilde{W}_i^r - w_i^{opt_2} = \sum_{k=1}^m \frac{\partial \Psi_i}{\partial x_k}(\sigma)(\tilde{\sigma}_k - \sigma_k) + \frac{1}{2}\sum_{k,j=1}^m \frac{\partial^2 \Psi_i}{\partial x_k \partial x_j}(\mathbf{c}) \times (\tilde{\sigma}_k - \sigma_k)(\tilde{\sigma}_j - \sigma_j). \tag{39}$$

Finally, let us consider the function $f : \mathbb{R}^2 \to \mathbb{R}$ by $f(\mathbf{x}) = x_2 - x_1^2$ then we have for some $\mathbf{a} \in \mathbb{R}^2$:

$$f(\mathbf{x}) = f(\mathbf{a}) + \frac{\partial f}{\partial x_1}(\mathbf{a})(x_1 - a_1) + \frac{\partial f}{\partial x_2}(\mathbf{a})(x_2 - a_2) + \frac{1}{2}\sum_{k,j=1}^{2} \frac{\partial^2 f}{\partial x_k x_j}(\mathbf{d})(x_k - a_k)(x_j - a_j) = f(\mathbf{a}) - 2a_1(x_1 - a_1) + (x_2 - a_2) - (x_1 - a_1)^2.$$

Then using the previous expression of f with $\mathbf{x} = (\hat{P}_{i,1}, \hat{P}_{i,2})$ and $a = (P_{i,1}, P_{i,2})$, we get:

$$\tilde{\sigma}_i - \sigma_i = -2P_{i,1}(\hat{P}_{i,1} - P_{i,1}) + \hat{P}_{i,2} - P_{i,2} + \frac{1}{2}\sum_{k,j=1}^{2} \frac{\partial^2 f}{\partial x_k x_j}(\mathbf{d})(\hat{P}_{i,k} - P_{i,k})(\hat{P}_{i,j} - P_{i,j}). \tag{40}$$

Moreover, we can notice that

$$\mathbb{P}(\tilde{K} \neq K) = \mathbb{P}(\exists i \in I, \tilde{\sigma}_i \neq \sigma_i) \leq \sum_{i=1}^{m} \left[\mathbb{P}(\hat{P}_{i,1} \neq P_{i,1}) + \mathbb{P}(\hat{P}_{i,2} \neq P_{i,2})\right].$$

We choose $x > P_{i,1}$ and $y < P_{i,1}$ so that:

$$\mathbb{P}(\hat{P}_{i,1} \neq P_{i,1}) = \mathbb{P}(\hat{P}_{i,1} \geq x) + \mathbb{P}(\hat{P}_{i,1} \leq y) = \mathbb{P}(\hat{P}_{i,1} \geq x) + \mathbb{P}(-\hat{P}_{i,1} \geq -y).$$

A similar result can be written for $\mathbb{P}(\hat{P}_{i,2} \neq P_{i,2})$, then we apply the Cramer large deviation inequality (Dombry 2005 [11]) and we get

$$\mathbb{P}(\tilde{K} \neq K) \leq C \exp(-dn) \tag{41}$$

for some constants $C, d \in \mathbb{R}$. Now, with these results, we can look for the bias of the estimator $\hat{P}_{f,r}^{ADS-2}$:

$$\begin{aligned}\mathbb{E}(\hat{P}_{f,r}^{ADS-2}) &= \mathbb{E}\big(\mathbb{E}(\hat{P}_{f,r}^{ADS-2}|\mathcal{F}_r^{\gamma_1})\big)\\ &= \mathbb{E}\Bigg[\mathbb{E}\Bigg(\sum_{i\in\tilde{K}} \frac{\rho_i}{(1-\epsilon)n_i + \epsilon\rho_i n} \times \sum_{j=1}^{(1-\epsilon)n_i+\epsilon\rho_i n} \xi(\mathbf{A}_j^i) + \sum_{i\in\tilde{K}^c} \frac{\rho_i}{N_i^r}\sum_{j=1}^{N_i^r} \xi(\mathbf{A}_j^i)\Big|\mathcal{F}_r^{\gamma_1}\Bigg)\Bigg]\\ &= \mathbb{E}\Bigg[\sum_{i\in\tilde{K}} \frac{\rho_i}{(1-\epsilon)n_i + \epsilon\rho_i n} \times\Bigg(\sum_{j=1}^{(1-\epsilon)n_i} \xi(\mathbf{A}_j^i) + \sum_{j=(1-\epsilon)n_i+1}^{(1-\epsilon)n_i+\epsilon\rho_i n} \xi(\mathbf{A}_j^i)\Bigg)\Bigg]\\ &\quad + \mathbb{E}\Bigg[\sum_{i\in\tilde{K}^c} \frac{\rho_i}{N_i^r}\sum_{j=1}^{N_i^r} \mathbb{E}(\xi(\mathbf{A}_j^i)|\mathcal{F}_r^{\gamma_1})\Bigg].\end{aligned}$$

The first expectation term of the previous result is denoted by J_1 and the second by J_2. Then

$$J_1 = \mathbb{E}\Bigg[\sum_{i\in\tilde{K}} \frac{(1-\epsilon)n_i}{(1-\epsilon)n_i + \epsilon\rho_i n}\rho_i \hat{P}_{i,1} + \sum_{i\in\tilde{K}} \frac{\rho_i}{(1-\epsilon)n_i + \epsilon\rho_i n} \sum_{j=(1-\epsilon)n_i+1}^{(1-\epsilon)n_i+\epsilon\rho_i n} \xi(\mathbf{A}_j^i)\Bigg].$$

Now, we decompose J_1 according to $\mathbb{1}_{\tilde{K}=K}$ and $\mathbb{1}_{\tilde{K}\neq K}$:

$$J_1 = \sum_{i\in K} \frac{(1-\epsilon)n_i}{(1-\epsilon)n_i + \epsilon\rho_i n}\rho_i \mathbb{E}\big[\mathbb{1}_{\tilde{K}=K}\hat{P}_{i,1}\big] + \sum_{i\in K} \frac{\rho_i}{(1-\epsilon)n_i + \epsilon\rho_i n} \times \sum_{j=(1-\epsilon)n_i+1}^{(1-\epsilon)n_i+\epsilon\rho_i n} \mathbb{E}\big[\mathbb{1}_{\tilde{K}=K}\xi(\mathbf{A}_j^i)\big] + o(\frac{1}{n}).$$

We use the Cauchy-Schwarz inequality, the fact that ξ is bounded and result (41), and we get:

$$J_1 = \sum_{i\in K} \rho_i P_{i,1} + o(\frac{1}{n}).$$

Moreover, we have:

$$J_2 = \mathbb{E}\Big[\sum_{i\in\tilde{K}^c}\frac{\rho_i}{N_i^r}\Big(\sum_{j=1}^{(1-\epsilon)n_i}\xi(\mathbf{A}_j^i)$$
$$+\sum_{j=(1-\epsilon)n_i+1}^{(1-\epsilon)(n_i+n_i^r)}P_{i,1}$$
$$+\sum_{j=(1-\epsilon)(n_i+n_i^r)+1}^{N_i^r}P_{i,1}\Big)\Big]$$
$$=\mathbb{E}\Big[\sum_{i\in\tilde{K}^c}\frac{\rho_i}{N_i^r}$$
$$\times\Big((1-\epsilon)n_i\hat{P}_{i,1}+(1-\epsilon)n_i^rP_{i,1}+\epsilon\rho_i nP_{i,1}\Big)\Big]$$
$$=\mathbb{E}\Big[\sum_{i\in\tilde{K}^c}\rho_iP_{i,1}\Big]$$
$$+\mathbb{E}\Big[\sum_{i\in\tilde{K}^c}\rho_i\frac{(1-\epsilon)n_i(\hat{P}_{i,1}-P_{i,1})}{N_i^r}\Big].$$

Now, as for J_1, we decompose J_2 according to $\mathbb{1}_{\tilde{K}^c=K^c}$ and $\mathbb{1}_{\tilde{K}^c\neq K^c}$, we use the Cauchy-Schwarz inequality, the fact that ξ is bounded and result (41), to get:

$$J_2=\sum_{i\in K^c}\rho_iP_{i,1}+\sum_{i\in K^c}\rho_i(1-\epsilon)J_3+o(\frac{1}{n})$$

with

$$J_3=\mathbb{E}\Big[\frac{n_i(\hat{P}_{i,1}-P_{i,1})}{N_i^r}\Big].$$

We use result (38) and substitute it into the expression of J_3:

$$J_3=-\frac{n_i}{n}\mathbb{E}\Big[(\hat{P}_{i,1}-P_{i,1})\Big(\frac{n}{N_i^{r,opt_2}}$$
$$+\frac{n^2(1-\epsilon)\gamma_2(\tilde{W}_i^r-w_i^{opt_2})}{(N_i^{r,opt_2})^2}$$
$$+\frac{g^{(2)}(c_1)}{2}(\tilde{W}_i^r-w_i^{opt_2})^2\Big)\Big]$$
$$=-\frac{n_in(1-\epsilon)\gamma_2}{(N_i^{r,opt_2})^2}\mathbb{E}\Big[(\hat{P}_{i,1}-P_{i,1})(\tilde{W}_i^r-w_i^{opt_2})\Big]$$
$$+\frac{n_i}{n}\mathbb{E}\Big[\frac{g^{(2)}(c_1)}{2}(\hat{P}_{i,1}-P_{i,1})(\tilde{W}_i^r-w_i^{opt_2})^2\Big].$$

Using the fact that $g^{(2)}$ is a continuous function and consequently bounded over any compact set, we can show that for some constant C:

$$\Big|\mathbb{E}\Big[\frac{g^{(2)}(c_1)}{2}(\hat{P}_{i,1}-P_{i,1})(\tilde{W}_i^r-w_i^{opt_2})^2\Big]\Big|\leq\frac{C}{n^{3/2}}.$$

So

$$J_3=-\frac{n_in(1-\epsilon)\gamma_2}{(N_i^{r,opt_2})^2}$$
$$\mathbb{E}\Big[(\hat{P}_{i,1}-P_{i,1})(\tilde{W}_i^r-w_i^{opt_2})\Big]+o(\frac{1}{n}).$$

We inject J_3 into J_2 and get the following expression of $\mathbb{E}(\hat{P}_{f,r}^{ADS-2})=J_1+J_2$:

$$\mathbb{E}(\hat{P}_{f,r}^{ADS-2})=P_f-\sum_{i\in K^c}\frac{\rho_in_in(1-\epsilon)^2\gamma_2}{(N_i^{r,opt})^2}$$
$$\times\mathbb{E}\Big[(\hat{P}_{i,1}-P_{i,1})(\tilde{W}_i^r-w_i^{opt_2})\Big]+o(\frac{1}{n})$$
$$=P_f-\sum_{i\in K^c}\frac{\rho_in_in(1-\epsilon)^2\gamma_2}{(N_i^{r,opt})^2}\times J_4+o(\frac{1}{n}),\tag{42}$$

with $J_4=\mathbb{E}\Big[(\hat{P}_{i,1}-P_{i,1})(\tilde{W}_i^r-w_i^{opt_2})\Big]$ and use equation (39) to get:

$$J_4=\sum_{k=1}^m\frac{\partial\Psi_i}{\partial x_k}(\sigma)\mathbb{E}\Big[(\hat{P}_{i,1}-P_{i,1})(\tilde{\sigma}_k-\sigma_k)\Big]$$
$$+\mathbb{E}\Big[\frac{1}{2}\sum_{k,j=1}^m\frac{\partial^2\Psi_i}{\partial x_k^2}(\mathbf{c})(\tilde{\sigma}_k-\sigma_k)(\tilde{\sigma}_j-\sigma_j)$$
$$\times(\hat{P}_{i,1}-P_{i,1})\Big].$$

As $\Psi_i\in C^3(\mathbb{R}^m_{+,*})$, we have for some constant C:

$$\Big|\mathbb{E}\Big[\frac{1}{2}\sum_{k,j=1}^m\frac{\partial^2\Psi_i}{\partial x_k\partial x_j}(\mathbf{c})(\tilde{\sigma}_k-\sigma_k)(\tilde{\sigma}_j-\sigma_j)$$
$$\times(\hat{P}_{i,1}-P_{i,1})\Big]\Big|\leq\frac{C}{n^{3/2}},$$

and as $(\hat{P}_{i,1}-P_{i,1})$ is centered and independent of $(\tilde{\sigma}_k-\sigma_k)$ for $k\neq i$:

$$J_4=\frac{\partial\Psi_i}{\partial x_i}(\sigma)\mathbb{E}\Big[(\hat{P}_{i,1}-P_{i,1})(\tilde{\sigma}_i-\sigma_i)\Big]+o(\frac{1}{n}).$$

Finally, we denote $J_5 = \mathbb{E}\Big[(\hat{P}_{i,1} - P_{i,1})(\tilde{\sigma}_i - \sigma_i)\Big]$ and get with (40):

$$J_5 = -2P_{i,1}\mathbb{E}\Big[(\hat{P}_{i,1} - P_{i,1})^2\Big] + \mathbb{E}\Big[(\hat{P}_{i,1} - P_{i,1}) \times (\hat{P}_{i,2} - P_{i,2})\Big] + \mathbb{E}\Big[\frac{1}{2}\sum_{k,j=1}^{2} \frac{\partial^2 f}{\partial x_k x_j}(\mathbf{d}) \times (\hat{P}_{i,k} - P_{i,k})(\hat{P}_{i,j} - P_{i,j})(\hat{P}_{i,1} - P_{i,1})\Big].$$

Once again, the last expectation of J_5 is $o(1/n)$, so we can rewrite it as:

$$J_5 = \frac{1}{n_i}\Big(P_{i,3} - P_{i,1}P_{i,2} - 2P_{i,1}^2(P_{i,2} - P_{i,1}^2)\Big) + o(\frac{1}{n}).$$

We substitute the estimations of J_5 into J_4 and J_4 into (42):

$$\mathbb{E}(\hat{P}_{f,r}^{ADS-2}) = P_f - \sum_{i\in K^c} \frac{\rho_i n(1-\epsilon)^2\gamma_2}{(N_i^{r,opt_2})^2}\frac{\partial \Psi_i}{\partial x_i}(\sigma) \times \Big(P_{i,3} - P_{i,1}P_{i,2} - 2P_{i,1}^2(P_{i,2} - P_{i,1}^2)\Big) + o(\frac{1}{n}).$$

As

$$\frac{\partial \Psi_i}{\partial x_i}(\sigma) = h_i(\gamma_1, \gamma_2, \sigma_{K^c}, \mathbf{w}_{K^c}, \rho_{K^c})$$

with $\sigma_{K^c} = (\sigma)_{i\in K^c}, \mathbf{w}_{K^c} = (w)_{i\in K^c}$ and $\rho_{K^c} = (\rho_i)_{i\in K^c}$ for some functions h_i, we find

$$\mathbb{E}(\hat{P}_{f,r}^{ADS-2}) = P_f - \sum_{i\in K^c} \frac{\rho_i n(1-\epsilon)^2\gamma_2}{(N_i^{r,opt_2})^2} \times \Big(P_{i,3} - P_{i,1}P_{i,2} - 2P_{i,1}^2(P_{i,2} - P_{i,1}^2)\Big) \times h_i(\gamma_1, \gamma_2, \sigma_{K^c}, \mathbf{w}_{K^c}, \rho_{K^c}) + o(\frac{1}{n})$$

which completes the proof. □

7.6 Proof of proposition 6

We assume that $\gamma_1(n) = \gamma_1$ and $\epsilon(n) = \epsilon$. The variance of the recycling estimator can be expressed as:

$$Var\big(\hat{P}_{f,r}^{ADS-2}\big) = \mathbb{E}\big(Var\big(\hat{P}_{f,r}^{ADS-2}|\mathcal{F}_r^{\gamma_1}\big)\big) + Var\big(\mathbb{E}\big(\hat{P}_{f,r}^{ADS-2}|\mathcal{F}_r^{\gamma_1}\big)\big).$$

We have

$$\hat{P}_{f,r}^{ADS-2} = \sum_{i\in\tilde{K}} \frac{\rho_i}{(1-\epsilon)n_i + \epsilon\rho_i n}\Big(\sum_{j=1}^{(1-\epsilon)n_i} \xi(\mathbf{A}_j^i) + \sum_{j=(1-\epsilon)n_i+1}^{(1-\epsilon)n_i+\epsilon\rho_i n} \xi(\mathbf{A}_j^i)\Big) + \sum_{i\in\tilde{K}^c} \frac{\rho_i}{N_i^r}\Big(\sum_{j=1}^{(1-\epsilon)n_i} \xi(\mathbf{A}_j^i) + \sum_{j=(1-\epsilon)n_i+1}^{N_i^r} \xi(\mathbf{A}_j^i)\Big). \tag{43}$$

As the two sums from $j = 1$ to $(1-\epsilon)n_i$ are known conditionally to $\mathcal{F}_r^{\gamma_1}$, we get using the variance properties:

$$\mathbb{E}\big(Var\big(\hat{P}_{f,r}^{ADS-2}|\mathcal{F}_r^{\gamma_1}\big)\big) = \mathbb{E}\Big[\sum_{i\in\tilde{K}} \frac{\rho_i^2\sigma_i^2\epsilon\rho_i n}{[(1-\epsilon)n_i + \epsilon\rho_i n]^2} + \sum_{i\in\tilde{K}^c} \frac{\rho_i^2\sigma_i^2}{(N_i^r)^2}\big((1-\epsilon)n_i^r + \epsilon\rho_i n\big)\Big].$$

We decompose the previous expectation according to $\mathbb{1}_{\tilde{K}=K}$ and $\mathbb{1}_{\tilde{K}\neq K}$, we use the Cauchy-Schwarz inequality, the fact that $\tilde{W}_i^r$ is bounded and result (41) to get:

$$\mathbb{E}\big(Var\big(\hat{P}_{f,r}^{ADS-2}|\mathcal{F}_r^{\gamma_1}\big)\big) = \frac{1}{n}\Big[\sum_{i\in K} \frac{\rho_i^2\sigma_i^2\epsilon\rho_i}{[(1-\epsilon)\gamma_1 w_i + \epsilon\rho_i]^2} + \sum_{i\in K^c} \rho_i^2\sigma_i^2 \times \mathbb{E}\Big(\frac{(1-\epsilon)\gamma_2\tilde{W}_i^r + \epsilon\rho_i}{[(1-\epsilon)(\gamma_1 w_i + \gamma_2\tilde{W}_i^r) + \epsilon\rho_i]^2}\Big)\Big] + o(\frac{1}{n}).$$

Now, let us get an estimation of

$$J = Var\big(\mathbb{E}\big(\hat{P}_{f,r}^{ADS-2}|\mathcal{F}_r^{\gamma_1}\big)\big).$$

Using expression (43) and the basics expectation properties, we have:

$$\mathbb{E}\big(\hat{P}_{f,r}^{ADS-2}|\mathcal{F}_r^{\gamma_1}\big) = P_f + \sum_{i\in\tilde{K}} \frac{\rho_i(1-\epsilon)n_i}{(1-\epsilon)n_i + \epsilon\rho_i n} \times \big(\hat{P}_{i,1} - P_{i,1}\big) + \sum_{i\in\tilde{K}^c} \frac{\rho_i(1-\epsilon)n_i}{N_i^r}\big(\hat{P}_{i,1} - P_{i,1}\big).$$

We decompose the previous expectation according to $\mathbb{1}_{\tilde{K}=K}$ and $\mathbb{1}_{\tilde{K}\neq K}$. We use the variance and covariance majoration by respectively the square mean and the L^2 scalar product, the Cauchy-Schwarz inequality, the fact that ξ and $\tilde{W}_i^r$ are bounded and result (41) and we get:

$$J = Var\Big[\sum_{i\in K} \frac{\rho_i(1-\epsilon)n_i}{(1-\epsilon)n_i+\epsilon\rho_i n}\big(\hat{P}_{i,1}-P_{i,1}\big) + \sum_{i\in K^c} \frac{\rho_i(1-\epsilon)n_i}{N_i^r}\big(\hat{P}_{i,1}-P_{i,1}\big)\Big] + o(\frac{1}{n}).$$

The first and the second terms of the previous variance are independent so that:

$$\begin{aligned} J &= Var\Big[\sum_{i\in K} \frac{\rho_i(1-\epsilon)n_i}{(1-\epsilon)n_i+\epsilon\rho_i n}\big(\hat{P}_{i,1}-P_{i,1}\big)\Big] \\ &+ Var\Big[\sum_{i\in K^c} \frac{\rho_i(1-\epsilon)n_i}{N_i^r}\big(\hat{P}_{i,1}-P_{i,1}\big)\Big] + o(\frac{1}{n}) \\ &= \frac{1}{n}\sum_{i\in K}\rho_i^2\sigma_i^2 \frac{(1-\epsilon)\gamma_1 w_i}{\big((1-\epsilon)\gamma_1 w_i+\epsilon\rho_i\big)^2} \\ &+ \sum_{i\in K^c}\rho_i^2\big((1-\epsilon)n_i\big)^2 Var\Big[\frac{\hat{P}_{i,1}}{N_i^r}\Big] + o(\frac{1}{n}). \end{aligned}$$

Let us denote $J_1 = Var\big[\hat{P}_{i,1}/N_i^r\big]$ and use result (38), thus:

$$\begin{aligned} J_1 &= \frac{\sigma_i^2}{(1-\epsilon)n_i(N_i^{r,opt_2})^2} + \frac{\big((1-\epsilon)\gamma_2 n\big)^2}{(N_i^{r,opt_2})^4} \\ &\times Var\big[(\tilde{W}_i^r - w_i^{opt_2})\hat{P}_{i,1}\big] \\ &+ \frac{1}{n^2} Var\big[C(\tilde{W}_i^r - w_i^{opt_2})^2\hat{P}_{i,1}\big] \end{aligned}$$

with C some bounded random variable. Once again, using the variance and covariance majoration by respectively the square mean and the L^2 scalar product, the Cauchy-Schwarz inequality and a Taylor expansion as (39) and (40) we get:

$$\big((1-\epsilon)n_i\big)^2 J_1 = \frac{1}{n}\frac{\sigma_i^2(1-\epsilon)\gamma_1 w_i}{\big((1-\epsilon)(\gamma_1 w_i+\gamma_2 w_i^{opt_2})+\epsilon\rho_i\big)^2} + o(\frac{1}{n}).$$

We inject the previous result into J and get:

$$\begin{aligned} J &= \frac{1}{n}\sum_{i\in K}\rho_i^2\sigma_i^2 \frac{(1-\epsilon)\gamma_1 w_i}{\big((1-\epsilon)\gamma_1 w_i+\epsilon\rho_i\big)^2} \\ &+ \frac{1}{n}\sum_{i\in K^c}\rho_i^2\sigma_i^2 \\ &\times \frac{(1-\epsilon)\gamma_1 w_i}{\big((1-\epsilon)(\gamma_1 w_i+\gamma_2 w_i^{opt_2})+\epsilon\rho_i\big)^2} \\ &+ o(\frac{1}{n}) \end{aligned}$$

and the variance expression of the recycling estimator becomes:

$$\begin{aligned} Var\big(\hat{P}_{f,r}^{ADS-2}\big) &= \frac{1}{n}\sum_{i\in K}\frac{\rho_i^2\sigma_i^2\epsilon\rho_i}{\big((1-\epsilon)\gamma_1 w_i+\epsilon\rho_i\big)^2} \\ &+ \frac{1}{n}\sum_{i\in K^c}\rho_i^2\sigma_i^2 \\ &\times \mathbb{E}\Big(\frac{(1-\epsilon)\gamma_2\tilde{W}_i^r+\epsilon\rho_i}{[(1-\epsilon)(\gamma_1 w_i+\gamma_2\tilde{W}_i^r)+\epsilon\rho_i]^2}\Big) \\ &+ \frac{1}{n}\sum_{i\in K}\rho_i^2\sigma_i^2\frac{(1-\epsilon)\gamma_1 w_i}{\big((1-\epsilon)\gamma_1 w_i+\epsilon\rho_i\big)^2} \\ &+ \frac{1}{n}\sum_{i\in K^c}\rho_i^2\sigma_i^2 \\ &\times \frac{(1-\epsilon)\gamma_1 w_i}{\big((1-\epsilon)(\gamma_1 w_i+\gamma_2 w_i^{opt_2})+\epsilon\rho_i\big)^2} \\ &+ o(\frac{1}{n}), \end{aligned}$$

which can be rewritten:

$$\begin{aligned} &Var\big(\hat{P}_{f,r}^{ADS-2}\big) \\ &= \frac{1}{n}\sum_{i\in K}\frac{\rho_i^2\sigma_i^2\epsilon\rho_i}{(1-\epsilon)\gamma_1 w_i+\epsilon\rho_i} \\ &+ \frac{1}{n}\sum_{i\in K^c}\rho_i^2\sigma_i^2\mathbb{E}\Big(\frac{(1-\epsilon)\gamma_2\tilde{W}_i^r+\epsilon\rho_i}{\big(N_i^r\big)^2}\Big) \\ &+ \frac{1}{n}\sum_{i\in K^c}\rho_i^2\sigma_i^2\frac{(1-\epsilon)\gamma_1 w_i}{(N_i^{r,opt_2})^2} + o(\frac{1}{n}). \end{aligned}$$

Once again using equation (38), we get $(1/N_i^r)^2 = (1/N_i^{r,opt_2})^2 + o(1/n^2)$ and finally:

$$Var\big(\hat{P}_{f,r}^{ADS-2}\big) = \frac{1}{n}\sum_{i\in K}\frac{\rho_i^2\sigma_i^2}{(1-\epsilon)\gamma_1 w_i + \epsilon\rho_i}$$
$$+\frac{1}{n}\sum_{i\in K^c}\rho_i^2\sigma_i^2$$
$$\times\frac{(1-\epsilon)\big(\gamma_1 w_i + \gamma_2\mathbb{E}(\tilde{W}_i^r)\big)+\epsilon\rho_i}{\bigg((1-\epsilon)\big(\gamma_1 w_i+\gamma_2 w_i^{opt_2}\big)+\epsilon\rho_i\bigg)^2}$$
$$+o(\frac{1}{n}),$$

which completes the proof. □

7.7 Proof of proposition 7

We need first to focus on algorithm 1 used to define $(\tilde{W}_i^r)_{i\in I}$, that we detail below.

(a) Compute, for $i = 1, ..., m$, *the quantities* $\frac{\beta_i}{\rho_i\sigma_i}$ *and sort them in decreasing order, denoting them by* $\frac{\beta_{(i)}}{\rho_{(i)}\sigma_{(i)}}$. *Compute, for* $i = 1, ..., m$, *the quantities*

$$\frac{1+\frac{\gamma_1(n)}{\gamma_2(n)}\sum_{j=i+1}^m\beta_{(j)}}{\frac{\gamma_1(n)}{\gamma_2(n)}\sum_{j=i+1}^m\rho_{(j)}\sigma_{(j)}}.$$

(b) Denote by i^* *the last* i *such as*

$$\frac{\beta_{(i)}}{\rho_{(i)}\sigma_{(i)}} \geq \frac{1+\dfrac{\gamma_1(n)}{\gamma_2(n)}\displaystyle\sum_{j=i+1}^m\beta_{(j)}}{\dfrac{\gamma_1(n)}{\gamma_2(n)}\displaystyle\sum_{j=i+1}^m\rho_{(j)}\sigma_{(j)}}.$$

If this inequality is false for all i, then, by convention, $i^* = 0$.

(c) Finally, for $i \leq i^*$, *set* $w_{(i)} = 0$, *and for* $i > i^*$, *set:*

$$w_{(i)} = \frac{\rho_{(i)}\sigma_{(i)}}{\displaystyle\sum_{j=i^*+1}^m\rho_{(i)}\sigma_{(i)}}\big(1+\frac{\gamma_1(n)}{\gamma_2(n)}\sum_{j=i^*+1}^m\beta_{(j)}\big) - \frac{\gamma_1(n)}{\gamma_2(n)}\beta_{(i)}$$

and turn back to the initial indexation.

$(\tilde{W}_i^r)_{i\in\{1,\ldots,m\}}$ is defined by replacing, in the previous algorithm, β_i by w_i for all i and σ_i by $\tilde{\sigma}_i$, for all i such that $\tilde{\sigma}_i > 0$, otherwise $\tilde{W}_i = 0$. Here, we suppose that all $(\tilde{\sigma}_i)_{i=1,\ldots,m}$ are strictly positive; if not, we just need to consider the positive ones, re-indexing them, and the following demonstration still stands.
Under the hypothesis $\gamma_1(n) \xrightarrow{n\to+\infty} 0$, we have $\gamma_1(n)/\gamma_2(n) \xrightarrow{n\to+\infty} 0$. There exists almost surely a random integer N such that for $n \geq N$ the inequality of point (b) of algorithm 1 will not be satisfied for any i. N must be chosen as the first integer such that:

$$\max_{i=1,\ldots,m}\left[\frac{w_{(i)}}{\rho_{(i)}\tilde{\sigma}_{(i)}}\sum_{j=i+1}^m\rho_{(j)}\tilde{\sigma}_{(j)}\right] < \frac{\gamma_2(N)}{\gamma_1(N)}.$$

Consequently, for $n \geq N$, we get $i^* = 0$ and

$$\tilde{W}_i^r = \frac{\rho_i\tilde{\sigma}_i}{\displaystyle\sum_{j=1}^m\rho_j\tilde{\sigma}_j}\big(1+\frac{\gamma_1(n)}{\gamma_2(n)}\big) - \frac{\gamma_1(n)}{\gamma_2(n)}w_i \geq 0 \tag{44}$$

for all $i \in I$.
Thus, for $n \geq N$, the estimator becomes:

$$\hat{P}_{f,r}^{ADS-2} = \sum_{i=1}^m\frac{\rho_i}{N_i^r}\sum_{j=1}^{N_i^r}\xi(\mathbf{A}_j^i)$$

with $N_i^r = (1-\epsilon(n))n\bar{W}_i^r + \epsilon(n)\rho_i n$ where $\bar{W}_i = \rho_i\tilde{\sigma}_i/\sum_{j=1}^m\rho_j\tilde{\sigma}_j$ and result (44) implies that:

$$(1-\epsilon(n))\bar{W}_i^r \geq (1-\epsilon(n))\gamma_1(n)w_i$$

for all i. We denote $N_i^{r,opt_1} = (1-\epsilon(n))nw_i^{opt_1} + \epsilon(n)\rho_i n$.
Now, for $n \geq N$, we have:

$$\sqrt{n}(\hat{P}_{f,r}^{ADS-2} - P_f) = \sqrt{n}\sum_{i=1}^m\frac{\rho_i}{N_i^r}\sum_{j=1}^{N_i^r}Y_{i,j}$$

with $Y_{i,j} = \xi(\mathbf{A}_i^j) - P_{i,1}$. Moreover

$$\sqrt{n}(\hat{P}_{f,r}^{ADS-2} - P_f) = \sum_{i=1}^m \frac{\sqrt{n}\rho_i}{N_i^r} \sum_{j=1}^{(1-\epsilon(n))\gamma_1(n)w_i n} Y_{i,j} + \sum_{i=1}^m \frac{\sqrt{n}\rho_i}{N_i^r} \sum_{j=(1-\epsilon(n))\gamma_1(n)w_i n+1}^{N_i^{r,opt_1}} Y_{i,j} + \bar{\epsilon}$$

with

$$\bar{\epsilon} = \sum_{i=1}^m \frac{\sqrt{n}\rho_i}{N_i^r}\left[\mathbb{1}_{\bar{W}_i > w_i^{opt_1}} \sum_{j=N_i^{r,opt_1}+1}^{N_i^r} Y_{i,j} - \mathbb{1}_{\bar{W}_i \leq w_i^{opt_1}} \sum_{j=N_i^r+1}^{N_i^{r,opt_1}} Y_{i,j}\right].$$

We can write:

$$\sqrt{n}(\hat{P}_{f,r}^{ADS-2} - P_f) = J_1 + J_2 + \bar{\epsilon}$$

with

$$J_1 = \sum_{i=1}^m \frac{\rho_i\sqrt{(1-\epsilon(n))\gamma_1(n)w_i}}{(1-\epsilon(n))\bar{W}_i^r + \epsilon(n)\rho_i} \times \frac{\sum_{j=1}^{(1-\epsilon(n))\gamma_1(n)nw_i} Y_{i,j}}{\sqrt{(1-\epsilon(n))\gamma_1(n)nw_i}},$$

$$J_2 = \sum_{i=1}^m \frac{\rho_i\sqrt{(1-\epsilon(n))(w_i^{opt_1} - \gamma_1(n)w_i) + \epsilon(n)\rho_i}}{(1-\epsilon(n))\bar{W}_i^r + \epsilon(n)\rho_i} \times \frac{\sum_{j=(1-\epsilon(n))\gamma_1(n)nw_i+1}^{N_i^{r,opt_1}} Y_{i,j}}{\sqrt{(1-\epsilon(n))(nw_i^{opt_1} - \gamma_1(n)nw_i) + \epsilon(n)\rho_i n}}$$

and $\bar{\epsilon}$ previously defined.
Now, we prove that J_1 and $\bar{\epsilon}$ converge to 0 in probability and

$$J_2 \xrightarrow[n\to+\infty]{\mathcal{L}} \mathcal{N}(0, \sigma^2_{opt_1}).$$

Then the Slutsky theorem will imply the final result.

(1) $J_1 \xrightarrow[n\to+\infty]{\mathbb{P}} 0$. We have:

$$J_1 = \sum_{i=1}^m \bar{\alpha}_{1,i}\bar{I}_{1_i}$$

with

$$\bar{\alpha}_{1,i} = \frac{\rho_i\sqrt{(1-\epsilon(n))\gamma_1(n)w_i}}{(1-\epsilon(n))\bar{W}_i^r + \epsilon(n)\rho_i}$$

and

$$\bar{I}_{1,i} = \frac{1}{\sqrt{(1-\epsilon(n))\gamma_1(n)nw_i}} \sum_{j=1}^{(1-\epsilon(n))\gamma_1(n)nw_i} Y_{i,j}$$

for all i.

Under the hypotheses $\gamma_1(n) \to 0$, $\epsilon(n) \to 0$ and as $\bar{W}_i \xrightarrow[n\to+\infty]{p.s} w_i^{opt_1}$ for all i, we have:

$$\bar{\alpha}_{1,i} \xrightarrow[n\to+\infty]{p.s \Rightarrow \mathbb{P}} 0$$

for all $i \in I$.
Using the characteristic function method and the fact that $\gamma_1(n)n \xrightarrow[n\to+\infty]{} +\infty$, we have:

$$\bar{I}_{1,i} \xrightarrow[n\to+\infty]{\mathcal{L}} \mathcal{N}(0, \sigma_i^2)$$

for all $i \in I$.
Finally, the Slutsky theorem implies that:

$$J_1 \xrightarrow[n\to+\infty]{\mathbb{P}} 0.$$

(2) $J_2 \xrightarrow[n\to+\infty]{\mathcal{L}} \mathcal{N}(0, \sigma^2_{opt_1})$. We have:

$$J_2 = \bar{\alpha}_2 \cdot \bar{I}_2^t$$

with $\bar{\alpha}_2$ the m-dimensional vector defined by:

$$\bar{\alpha}_{2,i} := \frac{\rho_i\sqrt{(1-\epsilon(n))(w_i^{opt_1} - \gamma_1(n)w_i) + \epsilon(n)\rho_i}}{(1-\epsilon(n))\bar{W}_i^r + \epsilon(n)\rho_i}$$

for all $i \in I$ and $\bar{I}_2$ the m-dimensional vector defined by:

$$\bar{I}_{2,i} := \frac{1}{\sqrt{(1-\epsilon(n))(nw_i^{opt_1} - \gamma_1(n)nw_i) + \epsilon(n)\rho_i n}} \times \sum_{j=(1-\epsilon(n))\gamma_1(n)nw_i+1}^{N_i^{r,opt_1}} Y_{i,j}$$

for all $i \in I$.
Again, under the hypotheses $\gamma_1(n) \to 0$, $\epsilon(n) \to 0$ and as $\bar{W}_i \xrightarrow[n\to+\infty]{p.s} w_i^{opt_1}$, we have:

$$\bar{\alpha}_2 \xrightarrow[n\to+\infty]{p.s} \alpha_2 \quad \Rightarrow \quad \bar{\alpha}_2 \xrightarrow[n\to+\infty]{\mathbb{P}} \alpha_2$$

with $\alpha_2 := (\frac{\rho_i}{\sqrt{w_i^{opt_1}}})_{i=1,\dots,m}$ a m-dimensional vector.
Furthermore, using the characteristic function method and the hypothesis $\gamma_1(n)/\gamma_2(n) \to 0$, we get:

$$\bar{I}_2 \xrightarrow[n\to+\infty]{\mathcal{L}} \mathcal{N}(0, \Sigma)$$

with Σ a $m \times m$ diagonal matrix with $\Sigma_{ii} = \sigma_i^2$ for all i.
We conclude with the Slutsky theorem:

$$\bar{\alpha}_2 \cdot \bar{I}_2^t \xrightarrow[n\to+\infty]{\mathcal{L}} \alpha_2 \cdot \mathcal{N}(0, \Sigma)$$

and

$$\alpha_2 \cdot \mathcal{N}(0, \Sigma) = \mathcal{N}(0, \alpha_2^t \Sigma \alpha_2) = \mathcal{N}(0, \sigma_{opt_1}^2).$$

(3) $\bar{\epsilon} \xrightarrow[n\to+\infty]{\mathbb{P}} 0$. We denote for all $i \in I$: $B_i = \{\tilde{W}_i > w_i^{opt_1}\}$. We define $[x]_+$ as equal to x if x is positive and equal to zero if not and $\wedge$ the minimum operator. Then, we have for all $i \in I$:

$$\mathbb{E}\Big(\Big|\frac{\mathbb{1}_{B_i}}{\sqrt{n}} \sum_{j=N_i^{r,opt_1}+1}^{N_i^r} Y_{i,j}\Big|\Big)^2$$

$$\leq \frac{1}{n}\mathbb{P}(B_i)\mathbb{E}\Big(\big(\sum_{j=N_i^{r,opt_1}+1}^{N_i^r \wedge N_i^{r,opt_1}} Y_{i,j}\big)^2\Big)$$

by the Cauchy-Schwarz inequality.
Moreover

$$Var\Big(\sum_{j=N_i^{r,opt_1}+1}^{N_i^r \wedge N_i^{r,opt_1}} Y_{i,j}\Big) = \mathbb{E}\Big[\Big(\sum_{j=N_i^{r,opt_1}+1}^{N_i^r \wedge N_i^{r,opt_1}} Y_{i,j}\Big)^2\Big] - \mathbb{E}\Big(\sum_{j=N_i^{r,opt_1}+1}^{N_i^r \wedge N_i^{r,opt_1}} Y_{i,j}\Big)^2 = \mathbb{E}\Big[\Big(\sum_{j=N_i^{r,opt_1}+1}^{N_i^r \wedge N_i^{r,opt_1}} Y_{i,j}\Big)^2\Big].$$

Indeed, conditioning over $\tilde{W}_i$, the second term is equal to zero since $\mathbb{E}(Y_{i,j}) = 0$. Thus:

$$\mathbb{E}\Big(\Big|\frac{\mathbb{1}_{B_i}}{\sqrt{n}} \sum_{j=N_i^{r,opt_1}+1}^{N_i^r \wedge N_i^{r,opt_1}} Y_{i,j}\Big|\Big)^2$$

$$\leq \frac{1}{n}\mathbb{P}(B_i) Var\Big(\sum_{j=N_i^{r,opt_1}+1}^{N_i^r \wedge N_i^{r,opt_1}} Y_{i,j}\Big)$$

$$\leq \frac{1}{n}\mathbb{P}(B_i)\mathbb{E}([N_i^r - N_i^{r,opt_1}]_+) Var(Y_i) + \mathbb{E}(Y_i)^2 Var([N_i^r \wedge N_i^{r,opt_1}]_+)$$

applying the second Wald identity

$$\leq \mathbb{P}(B_i)(1-\epsilon(n))\mathbb{E}\big([(\bar{W}_i - w_i^{opt_1}]_+)\sigma_i^2$$

since $\mathbb{E}(Y_i) = 0$

$$\leq \mathbb{P}(B_i)\mathbb{E}\big(|\bar{W}_i - w_i^{opt_1}|\big)\sigma_i^2.$$

As $\bar{W}_i - w_i^{opt_1}$ converges almost surely to 0 and is bounded by 1 for all i, we can use the dominated convergence theorem, which implies that:

$$\frac{\mathbb{1}_{B_i}}{\sqrt{n}} \sum_{j=N_i^{r,opt_1}+1}^{N_i^r} Y_{i,j} \xrightarrow[n\to+\infty]{L^1} 0. \tag{45}$$

In the same way, we demonstrate that for all $i \in I$:

$$\frac{\mathbb{1}_{B_i^c}}{\sqrt{n}} \sum_{j=N_i^r+1}^{N_i^{r,opt_1}} Y_{i,j} \xrightarrow[n\to+\infty]{L^1} 0. \tag{46}$$

It is also clear that for all $i \in I$:

$$\frac{\rho_i}{(1-\epsilon(n))\bar{W}_i + \epsilon(n)\rho_i} \xrightarrow[n\to+\infty]{p.s} \frac{\rho_i}{w_i^{opt_1}}. \tag{47}$$

The almost sure convergence and L^1 convergence both imply the convergence in probability. Therefore, convergence results (45), (46) and (47) stand in probability and we have:

$$\bar{\epsilon} \xrightarrow[n\to+\infty]{\mathbb{P}} 0,$$

which completes the proof of the first asymptotic result.

We remind that we denote

$$N_i^{r,opt_2} = (1-\epsilon)\big(n_i + n_i^{opt_2}\big) + \epsilon\rho_i n$$

with $n_i^{opt_2} = \gamma_2(n) n w_i^{opt_2}$ and $n_i = \gamma_1(n) n w_i$. Now we suppose $\gamma_1(n) = \gamma_1$. We have:

$$\sqrt{n}(\hat{P}_{f,r}^{ADS-2} - P_f) = \sum_{i=1}^{m} \frac{\sqrt{n}\rho_i}{N_i^r} \sum_{j=1}^{N_i^r} Y_{i,j}$$
$$= \sum_{i \in \tilde{K}} \frac{\sqrt{n}\rho_i}{N_i^r} \sum_{j=1}^{N_i^r} Y_{i,j} + \sum_{i \in \tilde{K}^c} \frac{\sqrt{n}\rho_i}{N_i^r} \sum_{j=1}^{N_i^r} Y_{i,j}$$

with $Y_{i,j} = \xi(\mathbf{A}_i^j) - P_i$. We can decompose the previous expression according to $\mathbb{1}_{\tilde{K}=K}$ and $\mathbb{1}_{\tilde{K}\neq K}$ and use the fact that ξ is bounded and result (41), to get:

$$\sqrt{n}(\hat{P}_{f,r}^{ADS-2} - P_f) = J + \bar{\epsilon}$$

with

$$\bar{\epsilon} \xrightarrow[n\to+\infty]{L^1} 0 \tag{48}$$

and

$$J = \sum_{i \in K} \frac{\sqrt{n}\rho_i}{N_i^r} \sum_{j=1}^{N_i^r} Y_{i,j} + \sum_{i \in K^c} \frac{\sqrt{n}\rho_i}{N_i^r} \sum_{j=1}^{N_i^r} Y_{i,j}.$$

Then,

$$J = \sum_{i \in K} \frac{\sqrt{n}\rho_i}{(1-\epsilon(n))n_i} \sum_{j=1}^{(1-\epsilon(n))n_i} Y_{i,j}$$
$$+ \sum_{i \in K^c} \frac{\sqrt{n}\rho_i}{N_i^r} \sum_{j=1}^{(1-\epsilon(n))n_i} Y_{i,j}$$
$$+ \sum_{i \in K^c} \frac{\sqrt{n}\rho_i}{N_i^r} \sum_{j=(1-\epsilon(n))n_i+1}^{N_i^r} Y_{i,j}$$
$$= \sum_{i \in K} \alpha_i \tilde{I}_i^1 + \sum_{i \in K^c} \tilde{\beta}_i \tilde{I}_i^1 + J_1$$

with

$$\tilde{I}_i^1 = \frac{1}{\sqrt{(1-\epsilon(n))n_i}} \sum_{j=1}^{(1-\epsilon(n))n_i} Y_{i,j} \text{ for } i \in I,$$

$$\alpha_i = \frac{\sqrt{n}\rho_i}{\sqrt{(1-\epsilon(n))n_i}} \text{ for } i \in K,$$

$$\tilde{\beta}_i = \frac{\sqrt{n}\rho_i \sqrt{(1-\epsilon(n))n_i}}{(1-\epsilon(n))(n_i + n_i^r) + \epsilon(n)\rho_i n} \text{ for } i \in K^c$$

and

$$J_1 = \sum_{i \in K^c} \frac{\sqrt{n}\rho_i}{N_i^r} \sum_{j=(1-\epsilon(n))n_i+1}^{N_i^r} Y_{i,j}.$$

J_1 can be expressed as:

$$J_1 = \sum_{i \in K^c} \frac{\sqrt{n}\rho_i}{N_i^r} \sum_{j=(1-\epsilon(n))n_i+1}^{N_i^{r,opt_2}} Y_{i,j} + \tilde{\epsilon}$$

with

$$\tilde{\epsilon} = \sum_{i \in K^c} \frac{\sqrt{n}\rho_i}{N_i^r} \Bigg[\mathbb{1}_{\tilde{W}_i^r > w_i^{opt_2}} \sum_{j=N_i^{r,opt_2}}^{N_i^r} Y_{i,j}$$
$$- \mathbb{1}_{\tilde{W}_i^r \leq w_i^{opt_2}} \sum_{j=N_i^r}^{N_i^{r,opt_2}} Y_{i,j} \Bigg].$$

Thus,

$$J_1 = \sum_{i \in K^c} \tilde{\delta}_i \tilde{I}_i^2 + \tilde{\epsilon}$$

with, for $i \in K^c$,

$$\tilde{\delta}_i = \frac{\sqrt{n}\rho_i \sqrt{N_i^{r,opt_2} - (1-\epsilon(n))n_i}}{N_i^r}$$

and

$$\tilde{I}_i^2 = \frac{1}{\sqrt{N_i^{r,opt_2} - (1-\epsilon(n))n_i}} \times \sum_{j=(1-\epsilon(n))n_i+1}^{N_i^{r,opt_2}} Y_{i,j}.$$

From all the previous arguments, we have:

$$\sqrt{n}(\hat{P}_{f,r}^{ADS-2} - P_f) = \sum_{i \in K} \alpha_i \tilde{I}_i^1 + \sum_{i \in K^c} \tilde{\beta}_i \tilde{P}_i^1$$
$$+ \sum_{i \in K^c} \tilde{\delta}_i \tilde{P}_i^2 + \bar{\epsilon} + \tilde{\epsilon}$$

and, as for the proof of the first asymptotical result, we can demonstrate that

$$\tilde{\epsilon} \xrightarrow[n\to+\infty]{L^1} 0. \tag{49}$$

Now, we denote $\alpha = (\alpha_i)_{i\in K}$, $\tilde{\beta} = (\tilde{\beta}_i)_{i\in K^c}$, $\tilde{\delta} = (\tilde{\delta}_i)_{i\in K^c}$, $\tilde{I}^1_K = (\tilde{P}^1_i)_{i\in K}$, $\tilde{I}^1_{K^c} = (\tilde{P}^1_i)_{i\in K^c}$, $\tilde{I}^2_{K^c} = (\tilde{P}^2_i)_{i\in K^c}$,

$$\beta = \left(\frac{\rho_i\sqrt{\gamma_1 w_i}}{w_i^{opt_2}}\right)_{i\in K^c}$$

and

$$\delta = \left(\frac{\rho_i\sqrt{w_i^{opt_2} - \gamma_1 w_i}}{w_i^{opt_2}}\right)_{i\in K^c}.$$

As, for all $i \in I$, $\tilde{W}^r_i \xrightarrow[n\to+\infty]{a.s} w_i^{opt_2}$, and as $\epsilon(n) \xrightarrow[n\to+\infty]{} 0$,

$$(\alpha, \tilde{\beta}, \tilde{\delta}) \xrightarrow[n\to+\infty]{a.s} (\alpha, \beta, \delta). \tag{50}$$

Also, using the characteristic function method, we obtain:

$$(\tilde{I}^1_K, \tilde{I}^1_{K^c}, \tilde{I}^2_{K^c}) \xrightarrow[n\to+\infty]{\mathcal{L}} \mathcal{N}(0, \Sigma) \tag{51}$$

with Σ a diagonal matrix with $(\sigma^2_K, \sigma^2_{K^c}, \sigma^2_{K^c})$ its diagonal vector where σ^2_K is the line vector $(\sigma^2_i)_{i\in K}$ and $\sigma^2_{K^c}$ the line vector $(\sigma^2_i)_{i\in K^c}$.
Using the Slutsky theorem and results (50) et (51), we get:

$$(\alpha, \tilde{\beta}, \tilde{\delta})\begin{pmatrix}\tilde{I}^1_K\\ \tilde{I}^1_{K^c}\\ \tilde{I}^2_{K^c}\end{pmatrix} \xrightarrow[n\to+\infty]{\mathcal{L}} \mathcal{N}(0, (\alpha,\beta,\delta)\Sigma(\alpha,\beta,\delta)^t) \tag{52}$$

or in other words:

$$\sum_{i\in K}\alpha_i\tilde{P}^1_i + \sum_{i\in K^c}\tilde{\beta}_i\tilde{P}^1_i + \sum_{i\in K^c}\tilde{\delta}_i\tilde{P}^2_i \xrightarrow[n\to+\infty]{\mathcal{L}} \mathcal{N}\left(0, \sigma^2_{opt_2} = \textstyle\sum_{i\in K}\frac{\rho_i^2\sigma_i^2}{\gamma_1 w_i} + \sum_{i\in K^c}\frac{\rho_i^2\sigma_i^2}{w_i^{opt_2}}\right).$$

We conclude applying the Slutsky theorem with results (48), (49) and (52), which completes the proof. $\square$

7.8 Proof of proposition 8

Let us begin with the consistency of the non-recycling estimator. The non-recycling estimator can be expressed as:

$$\hat{P}^{ADS-2}_{f,nr} = \sum_{i=1}^m \rho_i \frac{N_i^{nr,opt_1}}{N_i^{nr}} \times \frac{1}{N_i^{nr,opt_1}}\sum_{j=1}^{N_i^{nr,opt_1}} \xi(\mathbf{A}^i_{n_i+j}) + \bar{\epsilon}$$

with

$$\bar{\epsilon} = \sum_{i=1}^m \frac{\rho_i}{N_i^{nr}}\left(\mathbb{1}_{\tilde{W}_i^{nr} > w_i^{opt_1}} \sum_{j=N_i^{nr,opt_1}+1}^{N_i^{nr}} \xi(\mathbf{A}^i_{n_i+j}) - \mathbb{1}_{\tilde{W}_i^{nr} \le w_i^{opt_1}} \sum_{j=N_i^{nr}+1}^{N_i^{nr,opt_1}} \xi(\mathbf{A}^i_{n_i+j})\right).$$

As ξ is bounded, we have for some constant C:

$$|\bar{\epsilon}| \le C\sum_{i=1}^m \frac{|N_i^{nr} - N_i^{nr,opt_1}|}{N_i^{nr}}.$$

From the strong LLN, we can get that:

$$N_i^{nr} \xrightarrow[n\to+\infty]{a.s} N_i^{nr,opt_1}$$

and

$$\frac{1}{N_i^{nr,opt_1}}\sum_{j=1}^{N_i^{nr,opt_1}} \xi(\mathbf{A}^i_{n_i+j}) \xrightarrow[n\to+\infty]{a.s} P_{i,1}.$$

Thus

$$|\bar{\epsilon}| \xrightarrow[n\to+\infty]{a.s} 0 \quad \text{and} \quad \hat{P}^{ADS-2}_{f,nr} \xrightarrow[n\to+\infty]{a.s} P_f.$$

The unbiased property is immediate using the standard expectation properties, conditioning by $\mathcal{F}^{\gamma_1}_{nr}$.
Now, let us determine the expression of the variance of the non-recycling estimator. We have:

$$Var\big(\hat{P}^{ADS-2}_{f,nr}\big) = \mathbb{E}\Big[Var\big(\hat{P}^{ADS-2}_{f,nr}|\mathcal{F}^{\gamma_1}_{nr}\big)\Big] + Var\Big[\mathbb{E}\big(\hat{P}^{ADS-2}_{f,nr}|\mathcal{F}^{\gamma_1}_{nr}\big)\Big].$$

On one hand, we show that:

$$Var\Big[\mathbb{E}\big(\hat{P}^{ADS-2}_{f,nr}|\mathcal{F}^{\gamma_1}_{nr}\big)\Big] = 0.$$

On the other hand, we get:

$$\mathbb{E}\Big[Var\big(\hat{P}^{ADS-2}_{f,nr}|\mathcal{F}^{\gamma_1}_{nr}\big)\Big]$$
$$=\mathbb{E}\Big[\sum_{i=1}^{m}\frac{\rho_i^2}{(N_i^{nr})^2}\sum_{j=1}^{N_i^{nr}}Var\big(\xi(\mathbf{A}^i_{n_i+j})\big)\Big]$$
$$=\sum_{i=1}^{m}\rho_i^2\sigma_i^2\mathbb{E}\big(\frac{1}{N_i^{nr}}\big).$$

Finally, replacing N_i^{nr} by its value, we get:

$$Var\big(\hat{P}^{ADS-2}_{f,nr}\big)=\frac{1}{n}\sum_{i=1}^{m}\rho_i^2\sigma_i^2$$
$$\times\mathbb{E}\bigg(\frac{1}{[1-\epsilon(n)]\gamma_2(n)\frac{\rho_i\tilde{\sigma}_i}{\sum_{j=1}^{m}\rho_j\tilde{\sigma}_j}+\epsilon(n)\rho_i}\bigg),$$

which completes the proof. □

7.9 Proof of proposition 9

This demonstration is nearly the same as the one of proposition 7, so we will only give the sketch of the proof.
Denoting $Y_{i,j}=\xi(\mathbf{A}^i_j)-P_i$, we have:

$$\sqrt{n}(\hat{P}^{ADS-2}_{f,nr}-P_f)=\sum_{i=1}^{m}\rho_i\frac{\sqrt{n}}{N_i^{nr}}\sum_{j=1}^{N_i^{nr}}Y_{i,j}.$$

If we denote $B_i=\{\tilde{W}_i^{nr}>w_i^{opt_1}\}$, then:

$$\sqrt{n}(\hat{P}^{ADS-2}_{f,nr}-P_f)=J+\tilde{\epsilon}$$

with

$$J=\sum_{i=1}^{m}\frac{\sqrt{n}\rho_i\sqrt{N_i^{nr,opt_1}}}{N_i^{nr}}$$
$$\times\Big(\frac{1}{\sqrt{N_i^{nr,opt_1}}}\sum_{j=1}^{N_i^{nr,opt_1}}Y_{i,j}\Big)$$

and

$$\tilde{\epsilon}=\sum_{i=1}^{m}\frac{\rho_i\sqrt{n}}{N_i^{nr}}\Big[\mathbb{1}_{B_i}\sum_{j=N_i^{nr,opt_1}+1}^{N_i^{nr}}Y_{i,j}$$
$$-\mathbb{1}_{B_i^c}\sum_{j=N_i^{nr}+1}^{N_i^{nr,opt_1}}Y_{i,j}\Big].$$

First case: $\gamma_1(n)\xrightarrow[n\to+\infty]{}0$.
Using the same technique as in the proof of proposition 7, we prove that:

$$J\xrightarrow[n\to+\infty]{\mathcal{L}}\mathcal{N}(0,\sigma^2_{opt_1})$$

and

$$\tilde{\epsilon}\xrightarrow[n\to+\infty]{\mathbb{P}}0.$$

Thus, by the Slutsky theorem:

$$\sqrt{n}(\hat{P}^{ADS-2}_{f,nr}-P_f)\xrightarrow[n\to+\infty]{\mathcal{L}}\mathcal{N}(0,\sigma^2_{opt_1}).$$

Second case: $\gamma_1(n)=\gamma_1$.
Once again, using the same technique as in the proof of proposition 7, we can prove that:

$$J\xrightarrow[n\to+\infty]{\mathcal{L}}\mathcal{N}(0,\frac{1}{1-\gamma_1}\sigma^2_{opt_1})$$

and

$$\tilde{\epsilon}\xrightarrow[n\to+\infty]{\mathbb{P}}0.$$

Thus, by the Slutsky theorem:

$$\sqrt{n}(\hat{P}^{ADS-2}_{f,nr}-P_f)\xrightarrow[n\to+\infty]{\mathcal{L}}\mathcal{N}(0,\frac{1}{1-\gamma_1}\sigma^2_{opt_1}),$$

which completes the proof. □

7.10 Proof of proposition 10

With the multivariate central limit theorem, we get for all $i\in I$:

$$\sqrt{n}\big[(\hat{P}_{i,1},\hat{P}_{i,2})-(P_{i,1},P_{i,2})\big]\xrightarrow[n\to+\infty]{\mathcal{L}}\mathcal{N}(0,\tfrac{1}{\gamma_1 w_i}Q_i)$$

with

$$Q_i=\begin{pmatrix}P_{i,2}-P_{i,1}^2 & P_{i,3}-P_{i,1}P_{i,2}\\ P_{i,3}-P_{i,1}P_{i,2} & P_{i,4}-P_{i,2}^2\end{pmatrix}.$$

Now, we apply the Delta Method with the function $f:\mathbb{R}^2\to\mathbb{R}$ defined by $f(\mathbf{x})=x_2-x_1$ and we get:

$$\sqrt{n}\big[\tilde{\sigma_i}^2-\sigma_i^2\big]\xrightarrow[n\to+\infty]{\mathcal{L}}\mathcal{N}(0,\delta_i^2)$$

with $\delta_i^2 = \frac{1}{\gamma_1 w_i}\big(P_{i,4} - 4P_{i,1}P_{i,3} + 8P_{i,1}^2 P_{i,2} - P_{i,2}^2 - 4P_{i,1}^4\big)$ and $P_{i,j} = \mathbb{E}\big(\xi^j(\mathbf{A}^i)\big)$. Then, with the characteristic function method and the independence of the σ_i, we prove that:

$$\sqrt{n}\big[\tilde{\sigma}^2 - \sigma^2\big] \xrightarrow[n\to+\infty]{\mathcal{L}} \mathcal{N}(0,\Sigma)$$

with $\sigma^2 = (\sigma_i^2)_{i=1,\dots,m}$, $\tilde{\sigma}^2 = (\tilde{\sigma}_i^2)_{i=1,\dots,m}$ and Σ the diagonal matrix with $(\delta_1^2,\dots,\delta_m^2)$ its diagonal vector.

Let us define the functions: $\phi_i(\mathbf{x}) = \frac{\rho_i\sqrt{x_i}}{\sum_{j=1}^m \rho_j\sqrt{x_j}}$, $\Psi(x) = \big(\phi_1(\mathbf{x}),\dots,\phi_m(\mathbf{x})\big)$, $\nabla\phi_i$ the gradient of ϕ_i and J_Ψ the Jacobian matrix of Ψ. Finally, applying once again the Delta Method, with either ϕ_i or Ψ, to the previous asymptotical result, we obtain the desired results, which completes the proof. □

References

1. Au, S., Beck, J.: Estimation of small failure probabilities in high dimensions by subset simulation. Probabilistic Engineering Mechanics **16**, 263–277 (2001)
2. Blatman, G., Sudret, B.: An adaptive algorithm to build up sparse polynomial chaos expansions for stochastic finite element analysis. Prob Eng Mech **25**, 183–197 (2010)
3. Blatman, G., Sudret, B.: Adaptive sparse polynomial chaos expansion based on least angle regression. J Comput Phys **230**, 2345–2367 (2011)
4. Bungartz, H., Dirnstorfer, S.: Multivariate quadrature on adaptive sparse grids. Computing **71**, 89–114 (1985)
5. Cannamela, C.: Apport des méthodes probabilistes dans la simulation du comportement sous irradiation du combustible à particules. Ph.D. thesis, University of Paris VII (2007)
6. Chan, J., Kroese, A.: Rare-event probability estimation with conditional monte carlo. Ann. Oper. Res. (2009)
7. Cochran, W.: Sampling techniques (third edition). Wiley (1977)
8. Crestaux, T., Le Maître, O., MArtinez, J.M.: Plynomial chaos expansion for sensitivity analysis. Reliab Eng Sys Safety **94**, 1161–1172 (2009)
9. Dean, T., Dupuis, P.: Splitting for rare event simulation: a large deviations approach to design and analysis. Stoch. Proc. Appl. **119 (2)**, 562–587 (2009)
10. Del Moral, P., Garnier, J.: Genealogical particle analysis of rare events. The Annals of Applied Probability **15 (4)**, 2496–2534 (2005)
11. Dombry, C.: Quelques applications de la théorie des grandes déviations. Ph.D. thesis, University of Claude Bernard - Lyon 1 (2005)
12. Fang, K.T., Li, R., Sudjianto, A.: Design and modeling for computer experiments. Chapman & Hall/CRC. (2006)
13. Fang, K.T., S.Kotz, Ng, K.: Symmetric Multivariate and Related Distributions. Monographs on statistics and applied probability (ed. D. R. Cox, D. V. Hinkley, D. Rubin and B.W. Silverman) Chapman and Hall, London, New York (1990)
14. Gerstner, T., Griebel, M.: Numerical integration using sparse grids. Numer. Algorithms **18**, 209–232 (1998)
15. Gerstner, T., Griebel, M.: Dimension-adaptive tensor-product quadrature. Computing **71(1)**, 65–87 (2003)
16. Gille-Genest, A.: Utilisation des méthodes numériques probabilistes dans les applications au domaine de fiabilite des structures. Ph.D. thesis, University of Paris VI (1999)
17. Helton, J.: Uncertainty and sensitivity analysis techniques for use in performance assessment for radioactive waste disposal. Reliability Engineering & System Safety **42, issues 2-3**, 327–47 (1993)
18. Helton, J., Davis, F., Johnson, J.: A comparison of uncertainty and sensitivity analysis results obtained with random and latin hypercube sampling. Reliability Engineering & System Safety **89, issue 3**, 305–30 (2005)
19. Homem-de-Mello, T., Rubinstein, R.: Estimation of rare event probabilities using cross-entropy. Proceedings of the 2002 winter simulation conference (2002)
20. Lagnoux-Renaudie, A.: A two-step branching splitting model under cost constraint for rare event analysis. J. Appl. Prob. **46**, 429–452 (2009)
21. Lapeyre, B., Pardoux, E., Sentis, R.: Introduction aux méthodes de Monte Carlo. Springer (1997)
22. L'Ecuyer, P., Demers, V., Tuffin, B.: Splitting for rare-event simulation. Proceedings of the 2006 winter simulation conference (2006)
23. L'Ecuyer, P., Demers, V., Tuffin, B.: Rare events, splitting, and quasi-monte carlo. ACM Transactions on Modeling and Computer Simulation **17 (2)** (2007)
24. Li, G., Wang, S.W., Georgopoulos, P., Schoendorf, J., Rabitz, H.: Random sampling-high dimensional model representation (rs-hdmr) and orthogonality of its different order component functions. J. Phys. Chem. **110(7)**, 2474–85 (2006)
25. Liu, P., Kiureghian, A.D.: Structural reliability under incomplete probability information **112**, 85–104 (1986)
26. Madsen, H., Ditlevsen, O.: Strutural reliability methods. Wiley (1996)
27. Madsen, H., Krenk, S., Lind, N.: Methods of structural safety. Odile Jacob (2000)
28. Munoz Zuniga, M.: Méthodes stochastiques pour l'estimation contrôlée de faibles probabilités sur des modèles physiques complexes. application au domaine nucléaire. Ph.D. thesis, University of Paris VII (2011)
29. Munoz Zuniga, M., Garnier, J., Lefebvre, Y.: Controlled estimation of the probability of rare event for

a complex physical model - examination of monotoneous variation models (2008)
30. Munoz Zuniga, M., Garnier, J., Remy, E., de Rocquigny, E.: Adaptative Directional Stratification: an adaptive directional simulation method in a stratified space (2010)
31. Pedregal, P.: Introduction to optimization. Springer Berlin (2004)
32. Rasmussen, C., Williams, C.: Gaussian processes for machine learning. MIT Press and Cambridge. (2006)
33. Ripley, B.: Stochastic simulation. New york: Wiley Series in Probability and Statistics (1987)
34. Rubinstein, R., Kroese, D.: Simulation and the Monte-Carlo Method (second edition). Wiley (2007)
35. Santner, T., Williams, B., Notz, W.: The Design and Analysis of Computer Experiments. Springer (1999)
36. Siegmund, D.: Importance sampling in the monte-carlo study of sequential tests. Anals stat **4**, 673–84 (1976)
37. Soize, C., Ghanem, R.: Physical systems with random uncertainties: Chaos representations with arbitrary probability measure. SIAM J sci Comput. **26(2)**, 395–410 (2004)
38. Sudret, B.: Global sensitivity analysis using polynomial chaos expansions. Reliab Eng Sys Safety **93**, 964–79 (2008)
39. Todor, R., Schwab, C.: Convergence rates for sparse chaos approximations of elliptic problems with stochastic coefficients. IMA J. Numer. Anal. **27**, 232–261 (2007)
40. Zhang, P.: Nonparametric importance sampling. J. Am. Stat. Assoc **91(435)**, 1245–53 (1996)