Comparing conservative estimations of failure probabilities using sequential designs of experiments in monotone frameworks

V. Moutoussamy & N. Bousquet & B. Iooss
Department of Industrial Risk Management
EDF Research and Development, Chatou, France

P. Rochet & F. Gamboa & T. Klein
Institut de Mathématiques de Toulouse
Université Paul Sabatier, Toulouse, France.

ABSTRACT: Numerous structural reliability studies deal with the problem of estimating a failure probability associated to the exceedance of a security level by Monte Carlo simulation approaches. In practice, engineering studies involve time-consuming computer codes that make the cost of such methods prohibitive. This paper deals with the problem of developing accelerated Monte Carlo methods for such codes when the code output is assumed to be monotone with respect to the stochastic inputs. A noticeable gain provided by monotony is a deterministic bounding of the failure probability, hence a conservative estimation. Several strategies for exploring the input space through designs of numerical experiments can be proposed, that involve sequential optimization of criteria. This article provides a description of this framework and the comparison of these strategies with more classical approaches with the help of toy examples and real case-study.

1 INTRODUCTION

Computing the probability of undesirable, unobserved events is a common task in structural reliability engineering. When dealing with major risks occuring with low probability, the lack of observed failure data often requires to use so-called computer deterministic functions (or *codes*) reproducing the phenomenon of interest. The simulation of their uncertain inputs, being modeled as random variables, allows to compute statistical estimators of the probability. Usually these complex objects can be described as time-consuming black boxes. Therefore exploring the configurations of inputs leading to failure, in a non-intrusive way, requires to run the code over a design of numerical experiments. Classical (quasi) Monte Carlo designs cannot be practically used to explore configurations linked to low probabilities since they require too high computational budgets. Therefore, numerous sampling techniques has been proposed in the literature to diminish the computational cost while ensuring the precision of estimations. Most of them are based on choosing sequentially the elements of the design, by maximizing an expected gain in in-

formation at each step.

We consider the particular case of monotone codes. Even without regularity assumptions on the code, it is possible to take advantage of monotonicity to provide deterministic bounds (and thus conservative estimation) for the probability of failure when described as the probability that the output exceeds some fixed limit. Furthermore, when the design is chosen stochastically, a statistical estimator of the probability can be computed in parallel. Sequential designs can be built to refine these bounds and reduce the variance of this estimator when it exists. This article aims at defining some criteria for carrying out such strategies and comparing the performances of the corresponding designs, in terms of precision and computational cost.

Two main methodologies are especially emphasized. Making no prior assumption on the frontier between safety and failure domain in the input space, the first approach considers the next element as the one that maximizes the minimal gain in information (*maximin strategy*). On the contrary, the second approach uses classification tools for updating sequentially a prior on the frontier (*classification-based strategy*). In each case, an optimization task is needed at each step of the sequential strategy, that requires specific algorithms. The comparison task besides involves Monte Carlo standard approaches and engineering techniques standing on optimization rather than sampling (FORM). They are conducted over a range of toy examples in various dimensional settings. The interest of the study is finally highlighted by treating real hydrodological and thermo-mechanical case-studies.

2 A MONOTONE STRUCTURAL RELIABILITY FRAMEWORK

In this study, one assumes $G : \mathbb{U} \subset \mathbb{R}^d \to \mathbb{R}$ is a monotonic numerical code which represents the physical comportment of a phenomenon. The input vector $\mathbf{X} = (X_1, \cdots, X_d)$ of G is considered as random with continuous joint density $f_{\mathbf{X}}$ with support $\mathbb{U}$. The failure probability to estimate is

$$p = \mathbb{P}(G(\mathbf{X}) \leq 0) = \int_{\mathbb{U}} \mathbb{1}_{\{G(x)\leq 0\}} f_{\mathbf{X}}(x)dx. \quad (1)$$

G is assumed to be monotone, i.e. :

$$\forall \mathbf{x} \in \mathbb{U}, \forall i \in [\![1, d]\!], \forall \epsilon \in \mathbb{R}_+^*, \exists s_i \in \{-1, 1\},$$

$$G(x_1, \ldots, x_i + s_i\epsilon, \ldots, x_d) \leq G(x_1, \ldots, x_d).$$

Moreover, it is assumed that each X_i are independent. Since any joint probability distribution function (pdf), as the product of marginal pdf, is increasingly bijective in continuous cases, it is enough to consider that $f_{\mathbf{X}}$ is the uniform density over $\mathbb{U} = [0, 1]^d$ (up to a transformation). Defining the partial order $\mathbf{x} \preceq \mathbf{y} \Leftrightarrow \forall i \in \{1, \ldots, d\}, x_i \leq y_i$, an immediate consequence is that p can be bounded using any set of numerical experiments.

Indeed, let $\bar{\mathbf{x}}_n = \{\mathbf{x}_1, \ldots, \mathbf{x}_n\}$ be distributed on $\mathbb{U}$ and evaluated by G. Then denote $\Xi_n^- = \{\mathbf{x} \in \bar{\mathbf{x}}_n : G(\mathbf{x}) \leq 0\}$ and $\Xi_n^+ = \{\mathbf{x} \in \bar{\mathbf{x}}_n : G(\mathbf{x}) > 0\}$. Considering

$$\mathbb{U}_n^- = \{\mathbf{x} \in \mathbb{U} : \exists \mathbf{y} \in \Xi_n^-; \mathbf{x} \preceq \mathbf{y}\} \quad (2)$$

$$\mathbb{U}_n^+ = \{\mathbf{x} \in \mathbb{U} : \exists \mathbf{y} \in \Xi_n^+; \mathbf{x} \succeq \mathbf{y}\}, \quad (3)$$

two exact (deterministic) bounds for p can be obtained:

$$p_n^- \leq p \leq p_n^+, \quad (4)$$

where

$$p_n^- = \mathbb{P}(\mathbf{X} \in \mathbb{U}_n^-), \;\; p_n^+ = 1 - \mathbb{P}(\mathbf{X} \in \mathbb{U}_n^+)$$

with $\mathbf{X}$ uniformly distributed on $\mathbb{U}_n$ (see for instance Fig. 1). After few calls to G, an initialization step is build such that $p_0^-, p_0^+ \in]0, 1[$ and $\mathbb{U}_0 \subsetneq \mathbb{U}$. Bousquet (2012) considered this framework and developed a stochastic one-step-ahead strategy based on a nested uniform sampling of the next point of the design. In parallel to the progressive bounds $p_1^-, \ldots, p_n^-$ and $p_1^+, \ldots, p_n^+$, he proposed a stastistical M-estimator of p, the variance of which being significantly lower than the variance $p(1-p)/n$ of the usual Monte Carlo estimator. However, the gain brought by this

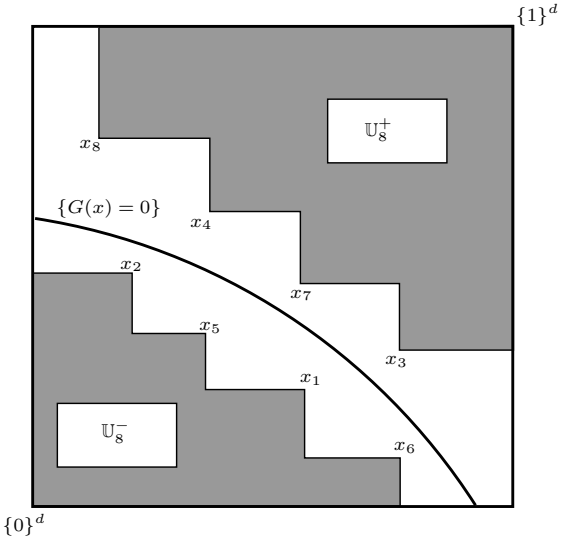

Figure 1: In dimension 2, $\Xi_8^- = \{\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_5, \mathbf{x}_6\}$ and $\Xi_8^+ = \{\mathbf{x}_3, \mathbf{x}_4, \mathbf{x}_7, \mathbf{x}_8\}$.

naive strategy can strongly diminish when the dimension increases and moreover, the bias affecting the estimator should be removed by bootstrap techniques in non-asymptotic settings. In the following, a sequential importance sampling approach is developed to improve this first approach, by contouring more accurately the failure surface

$$\Gamma = \{\mathbf{x} \in \mathbb{U} : G(\mathbf{x}) = 0\}. \tag{5}$$

3 SEQUENTIAL IMPORTANCE SAMPLING

Assume that at step n of one-step-ahead exploration of input space $\mathbb{U}$, the next point $\mathbf{x}_n$ of the design, at which G is computed, is sampled from the importance distribution

$$\mathbf{x}_n \sim f_{n-1} \equiv \mathcal{N}_d(\mathbf{x}^*, \sigma^2 I_d)\mathbb{1}_{\{\mathbf{x}\in\mathbb{U}_{n-1}\}} \tag{6}$$

where $\mathbb{U}_n = \mathbb{U} \setminus (\mathbb{U}_n^- \cup \mathbb{U}_n^+)$. The idea is to calibrate the distribution such that $\mathbf{x}^*$ be near Γ and the following statistical estimators present good convergence properties. A first (conditionally) unbiased estimator is

$$\bar{p}_n = (\bar{p}_n^- + \bar{p}_n^+)/2 \tag{7}$$

where, denoting $\xi_{\mathbf{x}} = \mathbb{1}_{\{G(\mathbf{x})\leq 0\}}$,

$$\bar{p}_n^- = p_{n-1}^- + \frac{\xi_{\mathbf{x}_n}}{f_{n-1}(\mathbf{x}_n)},$$

$$\bar{p}_n^+ = p_{n-1}^+ - \frac{1-\xi_{\mathbf{x}_n}}{f_{n-1}(\mathbf{x}_n)}.$$

and a conditional variance equal to

$$\frac{1}{4}\left(\mathbb{E}\left[\frac{1}{f_{n-1}^2(\mathbf{x}_n)}|\mathcal{F}_{n-1}\right] - \alpha_n^2\right) \tag{8}$$

with $\alpha_n = 2p - p_{n-1}^+ - p_{n-1}^-$. From conditional Jensen inequality the variance is greater than $(p_{n-1}^+ - p)(p - p_{n-1}^-) > 0$.

Considering deterministic weights $\omega = (\omega_1, \ldots, \omega_n) \in [0, n]^n$ such that $\sum_{k=1}^n \omega_k = n$, then a sequential importance sampling (unbiased) estimator of p is

$$\hat{p}_n = \frac{1}{n}\sum_{k=1}^n \omega_k \bar{p}_k$$

with optimized weights

$$\omega_k = n\frac{\mathbb{E}^{-1}\left[\mathrm{Var}\left[\bar{p}_k|\mathcal{F}_{k-1}\right]\right]}{\sum_{i=1}^n \mathbb{E}^{-1}\left[\mathrm{Var}\left[\bar{p}_i|\mathcal{F}_{i-1}\right]\right]}$$

$$= n\frac{\mathbb{E}^{-1}\left[\frac{1}{f_{k-1}^2(\mathbf{x}_k)} - \alpha_k^2\right]}{\sum_{i=1}^n \mathbb{E}^{-1}\left[\frac{1}{f_{i-1}^2(\mathbf{x}_i)} - \alpha_i^2\right]}$$

Now, we must notice that $\hat{p}_n$ is in the convexe hull of $\{\bar{p}_1, \ldots, \bar{p}_n\}$, i.e. $\hat{p}_n \in [\min\{\bar{p}_1, \ldots, \bar{p}_n\}, \max\{\bar{p}_1, \ldots, \bar{p}_n\}]$. That implies $\bar{p}_n$ cannnot be in $]p_n^-, p_n^+[$ with probability 1.

Denote $\|.\|$ the euclidian norm in $\mathbb{R}^d$. Choosing f_{n-1} as in (6) it can be rewrite like

$$f_{n-1}(\mathbf{x}) = \frac{e^{-\|\mathbf{x}-\mathbf{x}^*\|^2/2\sigma^2}}{\int_{\mathbb{U}_{n-1}} e^{-\|\mathbf{x}-\mathbf{x}^*\|^2/2\sigma^2} dx},$$

$$= \frac{e^{-\|\mathbf{x}-\mathbf{x}^*\|^2/2\sigma^2}}{\mathbb{E}\left[e^{-\|\mathbf{u}-\mathbf{x}^*\|^2/2\sigma^2}\right](p_{n-1}^+ - p_{n-1}^-)},$$

according to $\mathbf{u}$ is uniformly distributed in $\mathbb{U}_{n-1}$. The probability $\mathbb{P}(\bar{p}_n \in [p^-_{n-1}, p^+_{n-1}])$ is equal to

$$\mathbb{P}\left(\|\mathbf{x}-\mathbf{x}^*\|^2 \leq -2\sigma^2 \log \mathbb{E}[e^{-\|\mathbf{u}-\mathbf{x}^*\|^2/2\sigma^2}]\right),$$

and goes to 1 as σ goes to 0. The last point holds since $\mathbb{E}\left[\|\mathbf{u}-\mathbf{x}^*\|^2\right]$ do not depends of σ. Hence, choosing σ too low implies $\bar{p}_n \in]p^-_n, p^+_n[$ and $\hat{p}_n$ do not approximate correctly p. Otherwise, if σ is too large, one can have a better approximation of p by $\hat{p}_n$ but the gain of information given by the knowledge of $\mathbf{x}^*$ will be reduce since one simulate far of this optimal point. It is necessary to come to a compromise in the choice of σ. In practice, one choose $\sigma^2 = p^+_{n-1} - p^-_{n-1}$.

The variance of $\hat{p}_n$ can be written as :

$$\text{Var}\left[\hat{p}_n\right] = \frac{1}{\sum_{k=1}^{n} \mathbb{E}^{-1}\left[\text{Var}\left[\bar{p}_k|\mathcal{F}_{k-1}\right]\right]}.$$

The quantities α_n and $\mathbb{E}\left[\frac{1}{f^2_{n-1}(\mathbf{x}_n)}|\mathcal{F}_{n-1}\right]$ can be estimate respectively by $2\hat{p}_{n-1} - p^+_{n-1} - p^-_{n-1}$ and crude Monte Carlo method.

We purpose now two kinds of deterministic strategies. In all cases, one wants to maximise the information obtains from the knowledge of the next point tested state. The first one will be based on maximin criterion and the second one use an approximation of Γ from a classifier. The next point $\mathbf{x}^*$ to evaluate is such that

$$\mathbf{x}^* = \arg\max_{\mathbf{x}\in\mathbb{U}_n} C(\mathbf{x}), \tag{9}$$

where C is a function to maximize. From the form of $\mathbb{U}_n$ it is difficult to evaluate all points in this set by some functions. Few methods are useful, simulated annealing or develop a criteria and use a BFGS method. To reduce the computation time one make the choice to get the next point $\mathbf{x}^*$ in $\bar{\mathbf{y}}_N = (\mathbf{y}_1, \ldots, \mathbf{y}_N)$, N random variable uniformly distributed on $\mathbb{U}_n$.

4 CRITERION

4.1 *maximin*

The first approach will be based on a maximin criteria construct from the contribution of the next point to reduce the bounds. Let

$$p^{\pm}_{n+1}(\mathbf{x}) = \mathbb{P}(U \in \mathbb{U}^{\pm}_{n+1}(\mathbf{x})),$$

where

$$\mathbb{U}^-_{n+1}(\mathbf{x}) = \{\mathbf{z}\in\mathbb{U} : \exists \mathbf{y} \in (\Xi^-_n \cup \mathbf{x}); \mathbf{z} \preceq \mathbf{y}\}$$

$$\mathbb{U}^+_{n+1}(\mathbf{x}) = \{\mathbf{z}\in\mathbb{U} : \exists \mathbf{y} \in (\Xi^+_n \cup \mathbf{x}); \mathbf{z} \succeq \mathbf{y}\}.$$

A first function D is definied as

$$D(\mathbf{x}) = min(p^-_{n+1}(\mathbf{x}) - p^-_n, p^+_n - p^+_{n+1}(\mathbf{x})). \tag{10}$$

Assume $D(\mathbf{x}) = p^-_{n+1}(\mathbf{x}) - p^-_n$, then one can suppose $G(\mathbf{x}) \leq 0$. To be near of Γ is equivalent to keep away from $\mathbb{U}^-_n$. Then, one purpose to maximise $D(\mathbf{x})$ as first criterion.

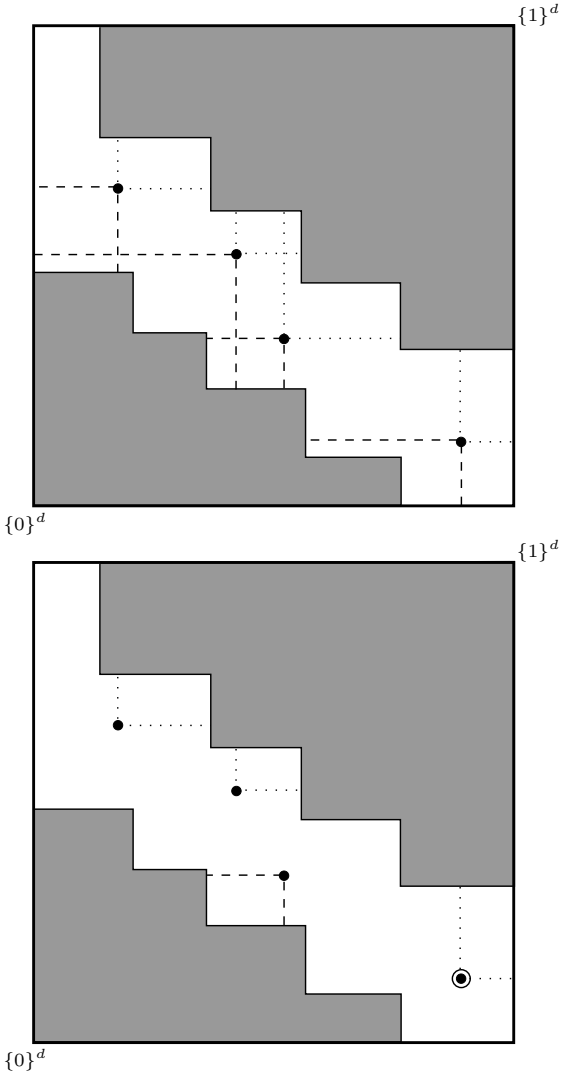

Figure 2: Illustration of the criterion D. Up : black points represents candidates to be $\mathbf{x}^*$. Dot (resp. dashed) lines delimite the volume contribution of upper (resp. lower) bound. Down : we keep the minimum of contribution for each candidates. The encircled point is choose as $\mathbf{x}^*$.

The computation of the bounds can be time consuming when the dimension is high. An alternative criterion $\widetilde{D}$ is proposed to accelerate the algorithm. Let

$$c^-_{n+1}(\mathbf{x}) = \#\{\mathbf{y} \in \bar{\mathbf{y}}_n : \mathbf{y} \preceq \mathbf{x}\}$$

$$c^+_{n+1}(\mathbf{x}) = \#\{\mathbf{y} \in \bar{\mathbf{y}}_n : \mathbf{y} \succeq \mathbf{x}\}.$$

where $\#A$ represent the number of elements in the set A.
Then,

$$\widetilde{D}(\mathbf{x}) = min(c^-_{n+1}(\mathbf{x}), c^+_{n+1}(\mathbf{x})). \quad (11)$$

When N is large enough it is clear that $\widetilde{D}$ is equivalent to D. Let C be D or $\widetilde{D}$, then our criterion is definied as

$$\mathbf{x}^* = \arg\max_{\mathbf{x} \in \bar{\mathbf{y}}_n} C(\mathbf{x}) \quad (12)$$

4.2 *Classification*

The second strategy is based on classification criteria. Since the output of G is binary, one can thinks to the k-nearest neighbor method. That is to class a point as failure or safe if it has more neighbors in the failure space or in the safety space. Two others tools commonly used are neural network and support vector machine (Hastie, Tibshirani, & Friedman 2008) . Neural networks are ued as our second criterion. Define $\pi_{-1}(\mathbf{x})$ (resp. $\pi_1(\mathbf{x})$) the weight given by neural network associed at $\mathbf{x}$ to be in the failure (resp. safety) space. Then one construct D :

$$D(\mathbf{x}) = [p^+_n - p^+_{n+1}(\mathbf{x})]\pi_1.$$

and

$$\mathbf{x}^* = \arg\max_{\mathbf{x} \in \bar{\mathbf{y}}_n} D(\mathbf{x}) \quad (13)$$

5 NUMERICALS STUDIES

A first toy exemple is defined as follow : in dimension d, let $\mathbf{X} = (X_1 \ldots, X_d)$ with $X_i \sim \Gamma(i+1, 1)$ and F_i the pdf of X_i. Denote :

$$Z_d = \frac{X_1}{\sum_{i=1}^{d} X_i} \sim \mathcal{B}eta(2, (d+1)(d+2)/2 - 3)$$

Let $q_{d,p}$ be the p-order quantile of Z_d, and define

$$G(\mathbf{X}) = Z_d - q_{d,p}.$$

Then,

$$p = \mathbb{P}(G(\mathbf{X}) \leq 0) = \mathbb{P}(Z_d \leq q_{d,p}).$$

The function G is increasing in his first direction and decreasing in the others.

In first, the comparison beetween the three differents strategies proposed in this paper is sumarized in table 1 : volume-maximin (V-Maximin), quick-maximin (Q-Maximin), classification(C-Maximin) tools and the one proposed by Bousquet(2012)

One presents the results obtains with FORM method and crude Monte Carlo with monotonic hypothesis (M-MC) in the table 2.

The estimator construct from Monte Carlo monotone is build as follow. Given say N calls to G, and let $\mathbf{x}$ be uniformly distributed in $\mathbb{U}$. Then, at step n, if $\mathbf{x}$ is in $\mathbb{U}_{n-1}^{-}$ or $\mathbb{U}_{n-1}^{+}$ it is not necesseray to test $G(\mathbf{x})$. Then the estimator of Monte Carlo method is given by

$$\hat{p}_{MC} = \frac{1}{M} \sum_{k \leq M} H(\mathbf{x}_k)$$

where

$$H(\mathbf{x}_k) = \mathbb{1}_{\{\mathbf{x}_k \in \mathbb{U}_{k-1}; G(\mathbf{x}_k) \leq 0\}} + \mathbb{1}_{\{\mathbf{x}_k \in \mathbb{U}_{k-1}^{-}\}},$$

and M represent the number of points such that one knows the state after N calls to G. It is clear that M is greater or equal to N. The second term of the right hand of the last equation do not depend of G, then a call of G is useless with probability $p_{k-1}^{+} - p_{k-1}^{-}$, and M can be see as a random variable since $\mathbf{x}$ is uniformly distributed in $\mathbb{U}$. Hence

$$\mathbb{P}(M > N) = 1 - \prod_{k=1}^{N} (p_k^{+} - p_k^{-}) \underset{N \to +\infty}{\longrightarrow} 1.$$

This last equation shows that Monte Carlo under monotony is more accurate, with high probability, than clasical Monte Carlo method without assumption.

Table 1: Comparison of criteria.

	$d=3$	$d=5$	$d=6$
	$p=10^{-4}$	$p=10^{-4}$	$p=10^{-3}$
Methods	$n=200$	$n=250$	$n=300$
MLE			
$\hat{p}_n$ $(\times p)$	1.20	0.99	1.07
p_n^{-} $(\times p)$	0.44	0.14	0.70
p_n^{+} $(\times p)$	3.68	14.7	24
CV(%)	12	14	18
Q-Maximin			
$\hat{p}_n$	3.072	9.1	11.7
p_n^{-}	0.43	0.02	0.02
p_n^{+}	4.88	15.4	22.9
CV(%)	15	14	14
C-Maximin			
$\hat{p}_n$	1.64	4.3	7.01
p_n^{-}	0.20	0.1	0.03
p_n^{+}	2.1	6.0	11.9
CV(%)	14	14	14

Table 2: Comparison of two classical methods.

	$d=3$	$d=5$	$d=6$
	$p=10^{-4}$	$p=10^{-3}$	$p=10^{-3}$
Methods	$n=200$	$n=250$	$n=300$
FORM			
$\hat{p}_n$ $(\times p)$	0.86	1.39	0.84
p_n^{-} $(\times p)$	0.26	0.16	0.08
p_n^{+} $(\times p)$	24	198	283
CV(%)	18	12	16
M-MC			
$\hat{p}_n$ $(\times p)$	1.76	2.28	0
p_n^{-} $(\times p)$	0.07	0.018	0
p_n^{+} $(\times p)$	24	163	313
CV(%)	46	70	∞
M	$n+25030$	$n+740$	$n+427$

About the bounds, the first method which uses a geometrical criteria seems to be equivalent to a uniform sample and do not implies a gain in information. The second one which make an evaluation of the failure surface seems to be really better and reduce significantly the width of the exact interval around p. The new estimator is not really good. This problem come from the construction of $\bar{p}_n$, a better way will be to choose a λ_n such that $\bar{p}_n = \lambda_n \bar{p}_n^{-} + (1 - \lambda_n)\bar{p}_n^{+}$.

6 CONCLUSIONS

In this paper we purposed two criteria to accelerate the convergence of deterministic bounds. The first one use geometrical property of the non dominated space, one sees in numerical examples that the gain is equivalent to the one proposed by Bousquet (2012). The second one approximates the failure surface by neural network. In practice, the use of these methods seems to reduce significantly the wide of the interval containing the failure probability. An other approach, not study here, is to create a meta-model $\hat{G}$ (e.g. gaussian process) no time consuming to remplace G and choose a point $\mathbf{x}$ such that $\hat{G}(\mathbf{x}) = 0$. An other possible way was to construct just one estimator whith a good convex linear combination of the bounds.

The quantities use to construct the estimator can be difficult to estimate. In particular the choice of a good candidate to be σ in the importance distribution. Here, one make a compromise with stay near of $\mathbf{x}^*$ and keep a low variance.

REFERENCES

Bousquet, N. (2012). Accelerated monte carlo estimation of exceedance probabilities under monotonicity constraints. *Annales de la Faculté des Sciences de Toulouse. Special Issue on Mathematical Methods for Design and Analysis of Numerical Experiments XXI*(3), 557–592.

De Rocquigny, E. (2009). Structural reliability under monotony : A review of properties of form and associated simulation methods and a new class of monotonous reliability methods (mrm). *Structural Safety 31*, 363–374.

Hastie, T., R. Tibshirani, & J. Friedman (2008). *The Elements of Statistical Learning*. Springer.

Kroese, D. & R. Rubinstein (2007). *Simulation and the Monte Carlo Method* (Wiley ed.).

Madsen, H. & O. Ditlevsen (1996). *Structural reliability methods* (Wiley ed.).