ANNALES
DE LA FACULTÉ
DES SCIENCES
DE
TOULOUSE
Mathématiques

NICOLAS BOUSQUET

Accelerated Monte Carlo estimation of exceedance probabilities under monotonicity constraints

Tome XXI, n° 3 (2012), p. 557-591.

<http://afst.cedram.org/item?id=AFST_2012_6_21_3_557_0>

Annales de la Faculté des Sciences de Toulouse Vol. XXI, n° 3, 2012
pp. 557–591

Accelerated Monte Carlo estimation of exceedance probabilities under monotonicity constraints

Nicolas Bousquet[(1)]

Abstract. — The problem of estimating the probability $p = P(g(\mathbf{X}) \leqslant 0)$ is considered when $\mathbf{X}$ represents a multivariate stochastic input of a monotonic function g. First, a heuristic method to bound p, originally proposed by de Rocquigny (2009), is formally described, involving a specialized design of numerical experiments. Then a statistical estimation of p is considered based on a sequential stochastic exploration of the input space. A maximum likelihood estimator of p build from successive dependent Bernoulli data is defined and its theoretical convergence properties are studied. Under intuitive or mild conditions, the estimation is faster and more robust than the traditional Monte Carlo approach, therefore adapted to time-consuming computer codes g. The main result of the paper is related to the variance of the estimator. It appears as a new baseline measure of efficiency under monotonicity constraints, which could play a similar role to the usual Monte Carlo estimator variance in unconstrained frameworks. Furthermore the bias of the estimator is shown to be corrigible via bootstrap heuristics. The behavior of the method is illustrated by numerical tests conducted on a class of toy examples and a more realistic hydraulic case-study.

Résumé. — On considère l'estimation de la probabilité $p = P(g(\mathbf{X}) \leqslant 0)$ où $\mathbf{X}$ est un vecteur aléatoire et g une fonction monotone. Premièrement, on rappelle et formalise une méthode, proposée par de Rocquigny (2009), permettant d'encadrer p par des bornes déterministes en fonction d'un plan d'expérience séquentiel. Le second et principal apport de l'article est la définition et l'étude d'un estimateur statistique de p tirant parti des bornes. Construit à partir de tirages uniformes successifs, cet estimateur présente sous de faibles conditions théoriques une variance asymptotique plus faible et une meilleure robustesse que l'estimateur classique de

(*) Reçu le 03/04/2011, accepté le 18/05/2012

(1) EDF Research & Development, Dpt. of Industrial Risk Management, 6 quai Watier, 78401 Chatou, France
nicolas.bousquet@edf.fr

Monte Carlo, ce qui rend la méthode adaptée à l'emploi de codes informatiques g lourds en temps de calcul. Des expérimentations numériques sont menées sur des exemples-jouets et un cas d'étude hydraulique plus réaliste. Une heuristique de boostrap, reposant sur un réplicat de l'hypersurface $\{\mathbf{x},\ g(\mathbf{x}) = 0\}$ par des réseaux de neurones, est proposée et testée avec succès pour ter le biais non-asymptotique de l'estimateur.

1. Introduction

In many technical areas, the exceedance of some unidimensional variable Z over a certain critical value z^* may define an event of probability p which has to be carefully estimated. Assumed to be stricly positive, p can be defined by

$$p \quad = \quad P(g(\mathbf{X}) \leqslant 0) \ = \ \int_{\mathbb{U}} \mathbb{1}_{\{g(\mathbf{x}) \leqslant 0\}} f(\mathbf{x})\ d\mathbf{x}$$

with $\mathbf{X}$ a random vector of uncertain input parameters with probability density function (pdf) f, taking its values in a $d-$dimensional space $\mathbb{U}$, and $g(\mathbf{X}) = z^* - Z$ a deterministic mapping from $\mathbb{U}$ to $\mathbb{R}$. This framework is often encountered in structural reliability studies [29], when g is a computer code reproducing a physical phenomenon. A Monte Carlo (MC) method is the usual way to estimate p by $\hat{p}_n = n^{-1} \sum_{k=1}^{n} \mathbb{1}_{\{g(\mathbf{x}_k) \leqslant 0\}}$ where n is large and the $\mathbf{x}_k$ are independently sampled from f. Avoiding regularity hypotheses on g, this unbiased estimator presents good convergence properties and an estimation error independent on d. Unfortunately, this strategy often appears inappropriate in practice when p reaches low values, since g can be time-consuming and the computational budget may be limited: a good estimation of a probability $p \sim 10^{-q}$ typically requires at least 10^{q+2} calls to g [25]. Furthermore, $\hat{p}_n$ has the theoretical defect not to be robust, in the sense given in [20]: its relative error, namely its coefficient of variation, does not tend to a finite limit when $p \to 0^+$, given any finite number n of trials.

Many non-intrusive strategies have been proposed to accelerate the MC approach. Traditional methods from the engineer community in structural reliability (FORM/SORM) treat the estimation of p as an optimization problem. The computational work is usually fast but the estimators suffer from weakly or non-controllable errors. Statistical approaches are judged in terms of reduction rate with respect to the MC estimator variance $\mathrm{Var}[\hat{p}_n] = p(1-p)/n$. Methods like quasi-MC, sequential MC or importance sampling [24] are based on selecting a *design of experiments* (DOE), namely a set of points in $\mathbb{U}$ on which g is tested, such that $\mathbb{U}$ be explored in areas close to

the limit state surface $\mathcal{S} = \{\mathbf{x} \in \mathbb{U} \ ; \ g(\mathbf{x}) = 0\}$. Most advanced methods often get rid of the time-consuming difficulties by emulating the behavior of g, for instance using kriging techniques [4] which presuppose smoothness conditions on g.

Minimizing the strength of regularity hypotheses placed on g underlies the development of specialized acceleration methods. For instance, computer codes can suffer from edge effects which restrict smoothness conditions [32]. On the other hand, the reality of the phenomenon can imply various form constraints on Z. Especially, the assumption that g is monotonic with respect to $\mathbf{x}$ is a standard problem in regression analysis [14]. In the area of numerical experiments, monotonicity properties of computer codes have been considered theoretically and practically, e.g. proving the MC acceleration of Latin Hypercube Sampling for the estimation of expectancies [28], carrying out screening methods for sensitivity analyses [27], constraining response surfaces [23, 22], predicting the behavior of network queuing systems [35], computing flood probabilities [13] or estimating the safety of a nuclear reactor pressure vessel [32].

Specific engineering works in structural reliability highlighted the possibility of bounding and estimating p significantly faster than using a MC approach. Under the name of *monotonic reliability methods* (MRM), de Rocquigny [13] proposed a class of algorithms contouring the limit state surface and enclosing p between deterministic bounds which dynamically narrow. A similar idea was explored in [34]. However, although a parallelization of such algorithms was already implemented [26], these methods were only empirically studied and some of the proposed estimators of p remained crude.

Therefore the present article aims at providing a first theoretical approach of the accelerated MC estimation of p when g is assumed to be monotonic and possibly discontinuous, although some smoothness constraints are assumed on the failure surface $\mathcal{S}$. More precisely, this article is structured as follows.

Section 2 is dedicated to a general description and a mathematical formalization of MRM. The main contribution is presented in Section 3: a statistical estimator of p is proposed, based on uniformly sampled DOEs in nested spaces. Defined as the maximum likelihood estimator of dependent Bernoulli data, its asymptotic properties are theoretically studied. The estimator is shown to be robust, and its variance gains a significant reduction with respect to the usual MC case. It may also be viewed as a baseline (or *target*) variance for monotonic structural reliability frameworks. The non-asymptotic bias of the estimator is examined in Section 4, through numerical experiments involving a class of toy examples. Based on a neural network

emulation of $\mathcal{S}$, bootstrap heuristics are proposed and successfully tested to remove this bias. Finally, a more realistic hydraulic case-study illustrates the benefits of the complete method.

Along the paper some connections are done with other areas of computational mathematics, especially about implementation issues, and a discussion section ends this article by focusing on the research avenues that must be explored in the area of stochastic sequential DOEs to improve the results presented here.

2. Material

2.1. Working assumptions, definitions and basic properties

Let $g : \mathbf{X} \mapsto g(\mathbf{X})$ be a deterministic function defined as a real-valued scalar mapping of $\mathbf{X} = (X_1, \ldots, X_d)$ on its definition domain $\mathbb{U} \subset I\!R^d$. *Deterministic* means that $g(\mathbf{x})$ produces always the same output if it is given the same input $\mathbf{x}$. Global monotonicity is defined as follows: $\forall i$, $\exists s_i \in \{-1, +1\}$, $\forall \epsilon > 0$, $\forall \mathbf{x} = (x_1, \ldots, x_d) \in \mathbb{U}$, such that

$$g\left(x_1, \ldots, x_{i-1}, x_i + s_i\epsilon, x_{i+1}, \ldots, x_d\right) \quad \leqslant \quad g\left(x_1, \ldots, x_{i-1}, x_i, x_{i+1}, \ldots, x_d\right)$$

where s_i represents the sign of monotonic dependence: $s_i = 1$ (*resp.* $s_i = -1$) when g is decreasing (*resp.* increasing) along with the $i-$th component x_i. The following assumption is made without loss of generality since any decreasing $i-$th component can be changed from x_i to $-x_i$:

Assumption 1. — The function g is globally increasing over $\mathbb{U}$.

To be general, $\mathbb{U} = [0, 1]^d$ and $\mathbf{X}$ is a random vector defined on the probability space $(\mathbb{U}, \mathcal{B}(\mathbb{U}), P)$. Next assumption is made following this same concern of generality.

Assumption 2. — All inputs $x_1, \ldots, x_d$ are independently uniform on $\mathbb{U} = [0, 1]^d$.

In real cases, $x_1, \ldots, x_d$ can be defined as *transformed inputs*, as usual in structural safety problems [29]. In such cases one can write $\mathbf{x} = T(\mathbf{y})$ where $\mathbf{y} = (y_1, \ldots, y_d)$ is a vector of *physical* inputs and T is their multivariate distributional transform. Therefore $g = \tilde{g} \circ T^{-1}$ where $\tilde{g}$ is a mononotic function and T has to preserve this monotonicity. When the y_i are independent, T is reduced to the vector of marginal cdfs $(F_1, \ldots, F_d)$ and is obviously increasing, so the assumption is not restrictive. Else, technical requirements

on T are needed, which depend on the way this joint distribution is defined [36]. See for instance [6] for such requirements on Gaussian copulas. Another general result is given in the Appendix (Supplementary Material).

Assumption 3. — Both subspaces $\mathbb{U}^- = \{\mathbf{x} \in \mathbb{U},\ g(\mathbf{x}) \leqslant 0\}$ and $\mathbb{U}^+ = \{\mathbf{x} \in \mathbb{U},\ g(\mathbf{x}) > 0\}$ are not empty (so that p exists in $]0,1[$).

Definition 2.1. — *A set of points of $\mathbb{U}$ is said to be safety-dominated (resp. failure-dominated) if g is guaranteed to be positive (resp. negative) in any point of this set.*

Denote by $\succeq$ the partial order between elements of $\mathbb{U}$ defined by $\mathbf{x} \succeq \mathbf{y} \Leftrightarrow x_k \geqslant y_k\ \forall k = 1, \ldots, d$. Then assume that some point value $g(\tilde{\mathbf{x}})$ is known, and consider the sets $\mathbb{U}^+_{\tilde{\mathbf{x}}} = \{\mathbf{x} \in \mathbb{U} \mid \mathbf{x} \succeq \tilde{\mathbf{x}}\}$ and $\mathbb{U}^-_{\tilde{\mathbf{x}}} = \{\mathbf{x} \in \mathbb{U} \mid \mathbf{x} \preceq \tilde{\mathbf{x}}\}$. The increasing monotonicity implies that if $g(\tilde{\mathbf{x}}) > 0$ (*resp.* $g(\tilde{\mathbf{x}}) < 0$), then $\mathbb{U}^+_{\tilde{\mathbf{x}}}$ is safety-dominated (*resp.* $\mathbb{U}^-_{\tilde{\mathbf{x}}}$ is failure-dominated). This proves next lemma.

Lemma 2.2. — *Both inequalities are true with probability 1:*

$$\begin{array}{rcll} p & \leqslant & 1 - P(\mathbf{X} \in \mathbb{U}^+_{\tilde{\mathbf{x}}}) & \text{if } g(\tilde{\mathbf{x}}) > 0, \\ p & \geqslant & P(\mathbf{X} \in \mathbb{U}^-_{\tilde{\mathbf{x}}}) & \text{else.} \end{array}$$

More generally, assume that n input vectors $(\mathbf{x}_j)_{j=1,\ldots,n}$ can be sorted into safe and failure sub-samples following the corresponding values of $\{g(\mathbf{x}_j)\}_{j=1,\ldots,n}$. They are respectively defined by

$$\Xi^+_n = \{\mathbf{x} \in (\mathbf{x}_j)_{j=1,\ldots,n} \mid g(\mathbf{x}_j) > 0\}$$

and

$$\Xi^-_n = \{\mathbf{x} \in (\mathbf{x}_j)_{j=1,\ldots,n} \mid g(\mathbf{x}_j) \leqslant 0\}.$$

Then one may define the sets

$$\begin{array}{rcl} \mathbb{U}^+_n & = & \left\{\mathbf{x} \in \mathbb{U} \mid \exists \mathbf{x}_j \in \Xi^+_n,\ \mathbf{x} \succeq \mathbf{x}_j\right\}, \\ \mathbb{U}^-_n & = & \left\{\mathbf{x} \in \mathbb{U} \mid \exists \mathbf{x}_j \in \Xi^-_n,\ \mathbf{x} \preceq \mathbf{x}_j\right\} \end{array}$$

(see Figure 1 for an illustration). Finally, denoting $p^-_n = P(\mathbf{X} \in \mathbb{U}^-_n)$ and $p^+_n = 1 - P(\mathbf{X} \in \mathbb{U}^+_n)$ to alleviate the notations, for all $n \geqslant 0$, in all the sequel one has

$$p^-_n \leqslant p \leqslant p^+_n. \tag{2.1}$$

Hereafter, $\mathbb{U}_n^+$ and $\mathbb{U}_n^-$ will be referred to as *dominated subspaces*, where the sign of $g(\mathbf{x})$ is known. Note that the complementary *non-dominated* subspace $\mathbb{U}_n = \mathbb{U}/\left(\mathbb{U}_n^+ \cup \mathbb{U}_n^-\right)$ is the only partition of $\mathbb{U}$ where further calls of g are required to improve the bounds. Finally, a topological assumption on $\mathcal{S}$ is needed to achieve the formal description of the situations studied in [13] and [26].

ASSUMPTION 4. — The limit state surface $\mathcal{S} = \{\mathbf{x} \in \mathbb{U} \ ; \ g(\mathbf{x}) = 0\}$ is regular enough and separates $\mathbb{U}$ in two disjoint domains $\mathbb{U}^-$ and $\mathbb{U}^+$ (*simply connected*).

The second part of this assumption implies that, in terms of classification, the two classes of points $\mathbb{U}_n^-$ and $\mathbb{U}_n^+$ are perfectly separable when $n \to \infty$. This property will be used later in the paper to carry out bootstrap heuristics. By *regular enough*, $\mathcal{S}$ is assumed not to be the surface of multidimensional stairs, so that it cannot be exhaustively described by a $n-$DOE with $n < \infty$. This mild assumption is ensured, for instance, if g is continuously differentiable on a non-empty measurable subset of $\mathcal{S}$. More formally, it is assumed that $\forall n < \infty$,

$$\sup_{\mathbf{x_n} \in \bar{\mathcal{S}}} \int_{\mathbb{U}_{n-1} \cap \mathbb{U}^-} \mathbb{1}_{\{\mathbf{x} \preceq \mathbf{x_n}\}} \, d\mathbf{x} \quad < \quad p - p_{n-1}^-, \tag{2.2}$$

$$\sup_{\mathbf{x_n} \in \bar{\mathcal{S}}} \int_{\mathbb{U}_{n-1} \cap \mathbb{U}^+} \mathbb{1}_{\{1-\mathbf{x} \preceq 1-\mathbf{x_n}\}} \, d\mathbf{x} \quad < \quad p_{n-1}^+ - p. \tag{2.3}$$

This will imply that $p_n^- < p < p_n^+$ and the finiteness of the strictly positive quantity $\tilde{\omega}_{n+1}(p) = [(p_n^+ - p)(p - p_n^-)]^{-1}$ encountered further in the paper.

Remark 2.3. — In multi-objective optimization, a dominated space can be interpreted as a subset of a performance space delimited by a Pareto frontier [18]. In this framework, g is thought as a monotonic rule of decision depending of d variables, for which the set of n best possible configurations (the frontier) is searched.

Remark 2.4. — The proportions $(p_n^-, 1 - p_n^+)$ are the volumes of two unions of hyperrectangles sharing the same orthogonal basis. Computing such volumes is known in computational geometry as Klee's measure problem, for which recursive sweepline algorithms can provide exact solutions [39]. Details about their implementation are given in Appendix (Supplementary Material). When d exceeds 4 or 5, these exact methods appear however too costly, and MC methods should be prefered in practice to compute these quantities.

2.2. MRM implementation: a one-step ahead strategy

Starting from $\mathbb{U}_0^+ = \{1^d\}$, $U_0^- = \{0^d\}$ and $\mathbb{U}_0 = \mathbb{U} = [0,1]^d$, the iterative scheme shared by all MRM variants at step $n \geqslant 1$ is based on:

---

1. selecting a DOE $\{\mathbf{x}_n^{(1)}, \ldots, \mathbf{x}_n^{(m_n)}\} \in \mathbb{U}_{n-1}$;
2. computing the *signatures* $\xi_{\mathbf{x_n}}^{(j)} = \mathbb{1}_{\left\{g\left(\mathbf{x_n^{(j)}}\right)<0\right\}}$;
3. updating the subspaces

$$\begin{aligned}
\mathbb{U}_n^- &= \mathbb{U}_{n-1}^- \cup \left\{\mathbf{x} \in \mathbb{U} \mid \exists\, \mathbf{x}_n^{(j)},\ \ \xi_{\mathbf{x_n}}^{(j)} = 1,\ \ \mathbf{x} \preceq \mathbf{x}_n^{(j)}\right\}, \\
\mathbb{U}_n^+ &= \mathbb{U}_{n-1}^+ \cup \left\{\mathbf{x} \in \mathbb{U} \mid \exists\, \mathbf{x}_n^{(j)},\ \ \xi_{\mathbf{x_n}}^{(j)} = 0,\ \ \mathbf{x} \succeq \mathbf{x}_n^{(j)}\right\}, \\
\mathbb{U}_n &= \mathbb{U}/(\mathbb{U}_n^- \cup \mathbb{U}_n^+)
\end{aligned}$$

4. updating the bounds $\{p_n^-, p_n^+\} = \{\mathrm{Vol}(\mathbb{U}_n^-), 1 - \mathrm{Vol}(\mathbb{U}_n^+)\}$.

Since $\mathbb{U}_n^- \subset \mathbb{U}_{n+1}^-$ $\forall n \geqslant 0$, then $P(\mathbf{X} \in \mathbb{U}_n^-) \leqslant P(\mathbf{X} \in \mathbb{U}_{n+1}^-)$ and the sequence (p_n^-) is nondecreasing. Symmetrically, the sequence (p_n^+) is nonincreasing. Since bounded in $[0,p]$ and $[p,1]$, both sequences are converging.

At each step, the DOE must be chosen accounting for the increasing monotonicity of g. Denoting $\mathbf{x_n^{(1)}}$ and $\mathbf{x_n^{(2)}}$ two elements of the DOE and assuming to know $\xi_{\mathbf{x_n}}^{(1)}$, it is unnecessary to compute $\xi_{\mathbf{x_n}}^{(2)}$ in two cases:

$$\begin{aligned}
&\text{if } \xi_{\mathbf{x_n}}^{(1)} = 1 \text{ and } \mathbf{x_n^{(1)}} \succeq \mathbf{x_n^{(2)}} \quad \Rightarrow \quad \mathbf{x_n^{(2)}} \in \mathbb{U}^-_{\mathbf{x_n^{(1)}}} \text{ and } \xi_{\mathbf{x_n}}^{(2)} = 1, \\
&\text{if } \xi_{\mathbf{x_n}}^{(1)} = 0 \text{ and } \mathbf{x_n^{(1)}} \preceq \mathbf{x_n^{(2)}} \quad \Rightarrow \quad \mathbf{x_n^{(2)}} \in \mathbb{U}^+_{\mathbf{x_n^{(1)}}} \text{ and } \xi_{\mathbf{x_n}}^{(2)} = 0.
\end{aligned}$$

Thus the order of trials should be carefully monitored, in relation with the partial order between the elements of the DOE. Reducing the DOE to a single element, i.e. $m_n = 1$ for all steps, minimizes the number of unnecessary trials. This one-step ahead strategy is favored in the present paper.

2.3. Stochastic MRM

Initialization. — First iterations should be monitored to reduce significantly the width of $[p_n^-, p_n^+]$, such that further iterations mainly focus on refinements. A deterministic strategy seems the most appropriate to start

from $[0,1]$ until providing non-trivial bounds. A dichotomic diagonal MRM, illustrated on Figure 2 in a two-dimensional case, was used in the examples considered further. It explores the non-dominated space in an intuitive way and stops at step $k_0 \geqslant 1$ such that

$$k_0 \quad \geqslant \quad 1 + \frac{\log(1/p)}{d \log 2}.$$

Consequently, an expected crude prior value of p can help to estimate the minimal number k_0 of trials. To alleviate the paper, the notation $(\mathbb{U}_0^+, \mathbb{U}_0^-, p_0^+, p_0^-)$ now describes the situation after $N-1$ introductive deterministic steps with $N \geqslant k_0 + 1$, such that $0 < p_0^- $ and $p_0^+ < 1$.

Switching to stochastic DOEs. — Pursuing a deterministic strategy can be too costly to be efficient, the upper bound p_n^+ offering possibly a very conservative assessment of p [13]. Intuitively, such a strategy should be optimized by selecting the next element of the DOE as the maximizer of a criterion which predicts a measure of dominated volume. Apart from the difficulty of predicting, choosing the criterion remains arbitrary. Switching to a stochastic strategy, which allows for a sequential statistical estimation of p in addition of providing bounds, seems a promising alternative approach. In this framework,

$$\mathbf{x_n} \quad \sim \quad f_{n-1}$$

at each step $n \geqslant 1$, with f_{n-1} a pdf defined on $\mathbb{U}_{n-1}$. Then the probability space $(\mathbb{U}, \mathcal{B}(\mathbb{U}), P)$ becomes endowed with the filtration $\mathbb{F} = (\mathcal{F}_n)$ where $\mathcal{F}_n$ is the $\sigma-$algebra generated by a $n-$sequence. The sequences $(p_0^-, \ldots, p_n^-)$ and $(p_0^+, \ldots, p_n^+)$ become monotonic, bounded stochastic processes with dependent increments.

Uniformly sampled DOEs. — The remainder of this article is devoted to a baseline statistical estimation of p in a monotonic framework, in a similar spirit to the MC approach in unconstrained frameworks. Therefore, in the following, the sampling is chosen uniform at each step: $\mathbf{x_n} \sim \mathcal{U}_{\mathbb{U}_{n-1}}$.

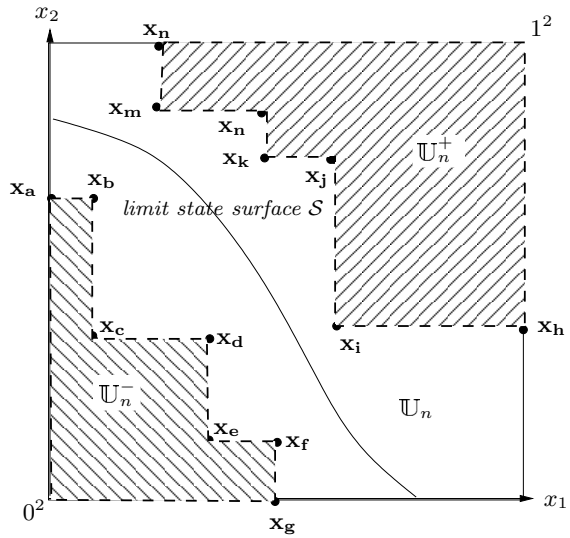

Figure 1. — Two-dimensional dominated and non-dominated subspaces after $n = 14$ iterations. Points $\{0^2, \mathbf{x_a}, \mathbf{x_b}, \mathbf{x_c}, \mathbf{x_d}, \mathbf{x_e}, \mathbf{x_f}, \mathbf{x_g}\}$ have nonzero signatures and are vertices of $\mathbb{U}_n^-$. Points $\{x_h, \mathbf{x_i}, \mathbf{x_j}, \mathbf{x_k}, \mathbf{x_l}, \mathbf{x_m}, \mathbf{x_n}, \mathbf{1^2}\}$ have zero signatures and are vertices of $\mathbb{U}_n^+$.

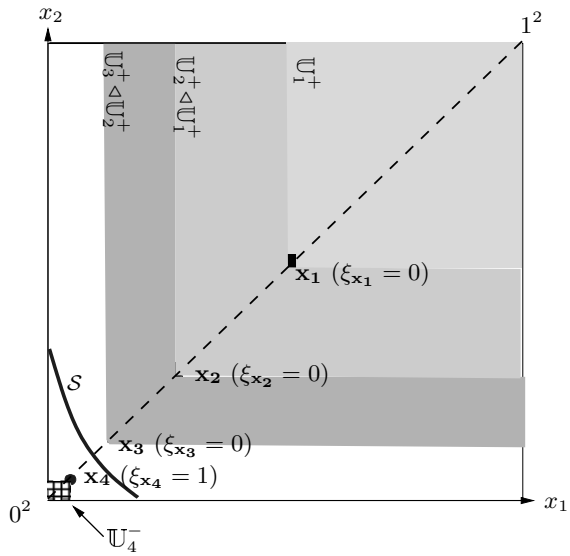

Figure 2. — Diagonal deterministic (DD-MRM) strategy, assuming a low p, stopping after 4 steps.

3. A maximum likelihood estimator of p

Assume that $\mathbf{x_1}, \ldots, \mathbf{x_n}$ are successively uniformly sampled in the nested non-dominated spaces $\mathbb{U}_0, \ldots, \mathbb{U}_{n-1}$. Next lemma follows.

LEMMA 3.1. — $p_n^-, p_n^+ \xrightarrow{a.s.} p$.

Therefore any estimator of p located between the bounds is strongly consistent. Especially, any crude average of the bounds gains a statistical validity. A more sophisticated approach can be carried out by noticing that, at step k, the occurence of a nonzero signature $\xi_{\mathbf{x_k}}$ follows a Bernoulli distribution $\mathcal{B}(\gamma_k)$ conditionally to $\mathcal{F}_{k-1}$, with

$$\begin{aligned} \gamma_k &= P\left(g(\mathbf{x}) \leqslant 0 | \mathbf{x} \in \mathbb{U}_{k-1}\right), \\ &= \frac{P\left(g(\mathbf{x}) \leqslant 0\right) - P\left(g(\mathbf{x}) \leqslant 0 | \mathbf{x} \in \mathbb{U}_{k-1}^-\right) P\left(\mathbf{x} \in \mathbb{U}_{k-1}^-\right)}{P\left(\mathbf{x} \in \mathbb{U}_{k-1}\right)} \end{aligned}$$

from Bayes' formula, hence

$$\gamma_k = \frac{p - p_{k-1}^-}{p_{k-1}^+ - p_{k-1}^-}. \tag{3.1}$$

After n steps, all information about p is brought by the dependent-data likelihood $L_n(p) = L_n(p|\mathbf{x_1}, \ldots, \mathbf{x_n})$ defined by the product of these conditional Bernoulli pdf:

$$L_n(p) = \prod_{k=1}^{n} \left(\frac{p - p_{k-1}^-}{p_{k-1}^+ - p_{k-1}^-}\right)^{\xi_{\mathbf{x_k}}} \left(\frac{p_{k-1}^+ - p}{p_{k-1}^+ - p_{k-1}^-}\right)^{1-\xi_{\mathbf{x_k}}}. \tag{3.2}$$

The associated maximum likelihood estimator (MLE) $\hat{p}_n$ is considered in next proposition.

PROPOSITION 3.2. — *Denote $\ell_n(p) = \log L_n(p)$. There exists a unique and consistent solution $\hat{p}_n$ in $]p_{n-1}^-, p_{n-1}^+[$ of the likelihood equation $\ell_n'(p) = \sum_{k=1}^n \tilde{\omega}_k(p)(p_k - p) = 0$, such that*

$$\hat{p}_n = \frac{\sum_{k=1}^{n} \tilde{\omega}_k(\hat{p}_n)\, p_k}{\sum_{k=1}^{n} \tilde{\omega}_k(\hat{p}_n)}, \tag{3.3}$$

with $\tilde{\omega}_k(p) = \left(\left(p - p_{k-1}^-\right)\left(p_{k-1}^+ - p\right)\right)^{-1}$ and $p_k = p_{k-1}^- + \left(p_{k-1}^+ - p_{k-1}^-\right)\xi_{\mathbf{x_k}}$

Assumption 4 ensures the existence of $\hat{p}_n$ since, by (2.2) and (2.3), p cannot be reached by at least one of the two bounds (p_{n-1}^-, p_{n+1}^-) for any finite n. Similarly, the quantities defined in next propositions remain finite if the limit state surface $\mathcal{S}$ has mild smoothness properties. They are related to the behavior of the inverse of the Fisher information associated to (3.2), which converges to 0 faster than the variance of the usual MC $n-$estimator

$$V_n^{MC}(p) \quad = \quad \frac{p(1-p)}{n}.$$

LEMMA 3.3. — *Assume that $\mathcal{S}$ is such that (2.2) and (2.3) hold (Assumption 4). Then, $\forall n \geqslant 0$,*

$$E\left[1/(p-p_n^-)^2\right] \quad < \quad \infty, \tag{3.4}$$
$$E\left[1/(p_n^+-p)^2\right] \quad < \quad \infty, \tag{3.5}$$

and consequently $E[\tilde{\omega}_{n+1}(p)] < \infty$.

PROPOSITION 3.4. — *Denote $J_n(p)$ the Fisher information associated to (3.2). Then*

$$J_n^{-1}(p) = \left(\sum_{k=1}^{n} E\left[\tilde{\omega}_k(p)\right]\right)^{-1} \leqslant V_n^{MC}(p) \frac{n}{\sum\limits_{k=1}^{n} (1-c_{k-1})^{-1}} < V_n^{MC}(p) \tag{3.6}$$

where $c_0 = 0$ and $\forall\ k > 1$,

$$c_k \quad = \quad E\left[\frac{p_k^-}{p} + \frac{1-p_k^+}{1-p} - \frac{p_k^-(1-p_k^+)}{p(1-p)}\right].$$

PROPOSITION 3.5. — *Denote $\gamma_0 = [(p_0^+ - p_0^-)/p_0^-]^2$. Then*

$$J_n^{-1}(p) \quad \leqslant \quad V_n^{MC}(p)\left(\frac{p\gamma_0}{1-p}\right). \tag{3.7}$$

In this data-dependent context, the central limit Theorem 3.6 remains classical in the sense that the Cramer-Rao bound given by the inverse of the Fisher information is still asymptotically reached by the MLE. This convergence is technically based on the martingality of the score process $n \mapsto \{\ell_n'(p)\}_n$. Therefore inequalities (3.6) and (3.7) imply asymptotic

variance reduction with respect to Monte Carlo and robustness. From (3.7), the asymptotic coefficient of variation (CV) of the MLE is such that

$$\mathrm{CV}\,[\hat{p}_n] \quad \leqslant \quad \frac{p}{\mathrm{E}\left[p_{n-1}^{-}\right]}\sqrt{\frac{\gamma_0}{n}} \;\infty\sim\; \sqrt{\frac{\gamma_0}{n}}.$$

THEOREM 3.6. — *Let* (λ_n) *be any deterministic sequence in* $]0,1[$ *such that* $\lambda_n \to 1$. *Under the supplementary assumptions:*

(i) $\dfrac{1}{n^\delta}\displaystyle\sum_{k=1}^{n}\left(\tilde{\omega}_k(p) - E[\tilde{\omega}_k(p)]\right) \xrightarrow{I\!P} 0 \quad$ *for any* $\delta \geqslant 1$.

(ii) $\dfrac{p_n^+ - p}{p - p_n^-} \xrightarrow{I\!P} 1.$

(iii) $\dfrac{\bar{p}_n - p}{p_n^+ - p} \xrightarrow{I\!P} 0 \quad$ *and* $\quad \dfrac{\bar{p}_n - p}{p - p_n^-} \xrightarrow{I\!P} 0 \quad$ *with* $\bar{p}_n = (1-\lambda_n)\hat{p}_n + \lambda_n p$

then

$$J_n^{1/2}(p)\,(\hat{p}_n - p) \xrightarrow{\mathcal{L}} \mathcal{N}(0,1). \tag{3.8}$$

The law of large numbers (i) reflects the requirement that the sum of weights $\tilde{\omega}_k(p)$ cannot diverge faster than $\mathcal{O}(J_n(p))$ from its mean behavior when $n \to \infty$. Although difficult to check in practice, this behavior seems rather intuitive because the sampling trajectories mainly vary at the first steps of the algorithm, when the non-dominated space is still large. Therefore (i) can be perceived as an indirect requirement on the shape of the surface $\mathcal{S}$. Assumption (ii) appears somewhat natural, saying that the bounds converge to p symmetrically. Assumption (iii) expresses the idea that any estimator located between $\hat{p}_n$ and p converges to p faster than the bounds. Again, it seems intuitive since $\hat{p}_n$ is defined as an incremental average (cf. (3.3)), and therefore adopts a smoother behavior than the bounds, as a function of n.

Next proposition allows for an empirical estimation of the asymptotic variance and confidence intervals. The additional requirement (v) appears mild and in the same spirit than the smoothness assumptions on $\mathcal{S}$, saying that p cannot be exactly reached by an average of the bounds for any finite number n of trials.

PROPOSITION 3.7. — *Denote* $\hat{J}_n(p) = \sum_{k=1}^{n} \tilde{\omega}_k(p)$. *Under the assumptions of Theorem 3.6, and assuming in addition:*

(iv) *Assumption* (i) *remains true* $\forall \delta \geqslant 1/2$,

(v) $\nexists\, n < \infty$ *such that* $p = (2n)^{-1} \sum_{k=1}^{n} \tilde{\omega}_k(p)(p_{k-1}^{-} + p_{k-1}^{-}) / \sum_{k=1}^{n} \tilde{\omega}_k(p)$,

then

$$\frac{\hat{J}_n^{\,5/2}(p)}{|\hat{J}_n'(p)|} \left(\hat{J}_n^{\,-1}(\hat{p}_n) - J_n^{-1}(p) \right) \quad \xrightarrow{\mathcal{L}} \quad \mathcal{N}(0,1). \qquad (3.9)$$

The reality of these theoretical descriptions is examined in the two next sections, through numerical experiments conducted on toy examples and a more realistic hydraulic model.

4. Numerical experiments I: toy examples

The statistical behavior of the MLE is illustrated here using the following generic toy example. For a given dimension d, denote

$$Z_d \quad = \quad h_d(\mathbf{Y}) \; = \; Y_1/(Y_1 + \sum_{i=2}^{d} Y_i)$$

where the physical input Y_i follows the gamma distribution $\mathcal{G}(i+1,1)$ with pdf F_{Y_i}, independently of other inputs. Obviously, $\forall\ d \geqslant 2$, h_d is increasing in $(-X_1, X_2 \ldots, X_d)$ where $X_i = F_{Y_i}(Y_i) \sim \mathcal{U}_{[0,1]}$, and Z_d follows the beta distribution $\mathcal{B}_e(2, 2^{-1}(d+1)(d+2) - 3)$. Therefore, denoting $q_{d,p}$ the $p-$order quantile of Y_d, the deterministic function defined by

$$g_d(\mathbf{X}) \quad = h_d \circ T^{-1}(\mathbf{X}) - q_{d,p},$$

with $T^{-1}(\mathbf{x}) = (F_{Y_1}^{-1}(x_1), \ldots, F_{Y_d}^{-1}(x_d))$, is related to the known exceedance probability p.

4.1. First results

In dimension 2, using $p = 5\%$, the behavior of MRM bounds can be easily compared to the MC 95%-confidence area (Figure 3). This small dimension induces a significant improvement in precision with respect to Monte Carlo, which however disappears in higher dimensions and highlights the need for

real statistical estimators. Studies of the root mean square error (RMSE) and the standard deviation of the MLE, which are plotted in Figure 5 as functions of the increasing number of calls to g_d for dimensions 3 and 4, have shown a high variance reduction of the iterative estimator $\hat{p}_n$ with respect to Monte Carlo but have highlighted a positive bias (Figure 4). Indeed the highest weights favor local estimators $p_k = p_k^+$ when approaching $\mathcal{S}$ (ie., when $\xi_{\mathbf{x_k}} = 1$ in (3.4)). On the examples considered in this last figure (as in other experiments not presented here), a marked gap in relative bias was noticed between dimensions 3 and 4. Under dimension 4, the bias remains reasonable from a moderate number of iterations (typically 400). Else it dramatically stays at high values. Other experiments have shown on this example the effective convergence of the empirical variance of the MLE towards the Cramer-Rao bound as well as the good behavior of its empirical estimate (Figure 6).

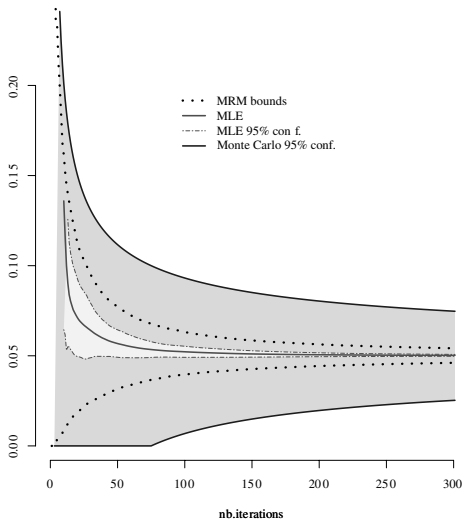

Figure 3. — MRM deterministic bounds and MLE, with Monte Carlo and MLE 95%-confidence areas, in dimension $d = 2$, for $p = 5\%$. Empirical estimations are made over 300 parallel MRM trajectories.

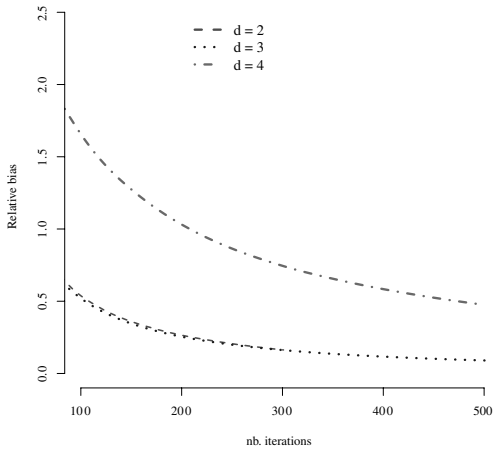

Figure 4. — Relative bias of the MLE $\hat{p}_n$ for the dimensions $d \in \{2, 3, 4\}$.

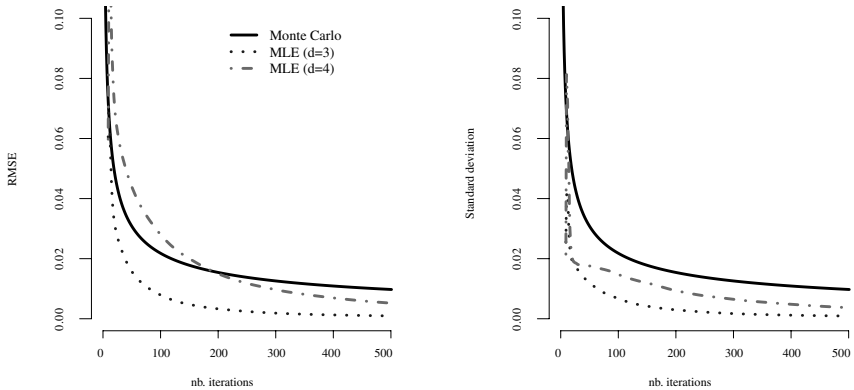

Figure 5. — Root mean square error (RMSE; *left*) and standard deviation (*right*) of the standard Monte Carlo estimator and $\hat{p}_n$ for $d = 3$ and $d = 4$, for $p = 0.05$. Empirical estimations are made over 300 parallel MRM trajectories.

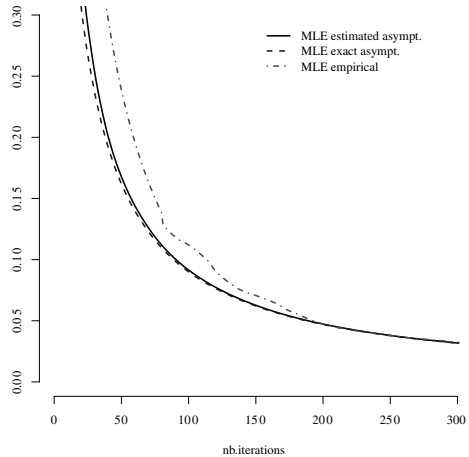

Figure 6. — Ratios of MLE standard deviations over Monte Carlo standard deviations, computed over 100 MRM replications, in dimension $d = 2$.

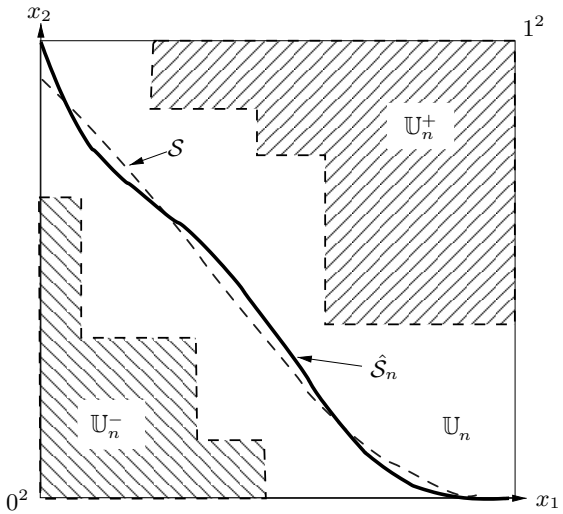

Figure 7. — Two-dimensional situation after $n = 14$ iterations. A replication $\hat{\mathcal{S}}_n$ (dashed curve) of $\mathcal{S}$ can be produced based on a monotonic neural network prediction of $\xi_{\mathbf{x}}$ in $\mathbb{U}_n$. The volume under $\hat{\mathcal{S}}_n$ corresponds to a new probability $\tilde{p}_n$ in the magnitude of p, which can be estimated by Monte Carlo at an arbitrary precision.

4.2. Bias correction via bootstrap heuristics

Bias removal appears as a practical requirement, automatizing the estimation of p. Indeed, given a finite value of n, estimating p requires to decide from which iteration $k_n \geqslant 1$ the computation of the MLE $\hat{p}_n$ can be worth it, redefining $\hat{p}_n = \sum_{i=k_n}^{n} \psi_{i,n}(\hat{p}_n) p_i$ with $\psi_{i,n}(p) = \tilde{\omega}_i(p) / \sum_{j=k_n}^{n} \tilde{\omega}_j(p)$. An intuitive rule is to select

$$k_n^* \quad = \quad \arg\min_{k_n} \mathrm{RMSE}\left(\hat{p}_n\right). \tag{4.1}$$

If the MLE were debiased, $\mathrm{RMSE}\left(\hat{p}_n\right) \simeq J_n^{-1}(p)$ which is minimized by $k_n^* = 1$.

Given a fixed number n of trials, two general approaches may be used for controlling and correcting the bias $B_n = \mathrm{E}[\hat{p}_n] - p$ affecting $\hat{p}_n$. A *corrective* approach consists in obtaining a closed-form expression for the bias from Taylor expansions [8, 17] or penalizing the score or the likelihood functions [3]. This approach is not carried out here since the data-dependent context would require a specific algebraic work and technical developments about the empirical estimation of the main quantities involved. The alternative use of *bootstrap* resampling techniques [15] can assess the bias empirically. For the simplicity of their principle, these heuristics are preferred here.

In the present context, bootstrap experiments must be based on a replication $\hat{\mathcal{S}}_n$ of the limit state surface $\mathcal{S}$ (see Figure 7 for an illustration). Under Assumption 4, $\mathcal{S}$ can be interpreted as the decision frontier of a supervised classification binary problem, without horseriding of classes (ie., perfectly separable). Therefore $\hat{\mathcal{S}}_n$ depends on the choice of a classifier $\hat{C}_{n,M}$ calibrated from an arbitrary number M of points sampled in dominated subspaces. The rationale of the bootstrap heuristics is as follows. Given $\hat{C}_{n,M}$, the signature $\xi_{\mathbf{x}}$ of any $\mathbf{x} \in \mathbb{U}_n$ can be predicted by the occurence of $P(g(\mathbf{x} \leqslant 0) | \hat{C}_{n,M}) \geqslant 1/2$. Then denote $\tilde{p}_{n,M}$ the volume under $\hat{\mathcal{S}}_n$. It can easily be estimated by $\tilde{p}_{n,M,Q}$ at an arbitrary precision by Monte Carlo sampling (depending on a number Q of particles). Moreover a large number S of MLE of $\tilde{p}_{n,M}$ can be fastly computed using the predicted signatures.

The bootstrap heuristics make sense if the classifier is chosen such that $\tilde{p}_{n,M} \to p$ when $(n, M) \to \infty$, so that the features of the experiment are asymptotically reproduced. Moreover, $\hat{C}_{n,M}$ must produce a monotonic surface $\hat{\mathcal{S}}_n$. For these reasons, the four-layer monotonic Multi-Layer neural networks (MLNN) proposed in [11] have been chosen for the experiments. Based on a combination of minimum and maximum functions (so-called MIN-MAX networks) over the two hidden layers, these networks have universal approximation capabilities of monotonic continuous functions. Besides, this choice

matches the advices by Hurtado [21] who strongly recommended MLNN and Support Vector Machines (SVM) to estimate $\mathcal{S}$ in a structural reliability context. Both tools are flexible, can estimate a frontier on the basis of a few samples and overcome the curse of dimensionality.

Classification-based bootstrap algorithm

---

1. Sample $\mathbf{x}^+ = (\mathbf{x}_\mathbf{1}^+, \ldots, \mathbf{x}_\mathbf{M}^+)\mathbf{iid}\sim\mathcal{U}_{\mathbb{U}_\mathbf{n}^+}$ and $\mathbf{x}^+ = (\mathbf{x}_\mathbf{1}^-, \ldots, \mathbf{x}_\mathbf{M}^-)\mathbf{iid}\sim\mathcal{U}_{\mathbb{U}_\mathbf{n}^-}$.
2. From $(\mathbf{x}^+, \mathbf{x}^-)$, build a monotonic classifier $\hat{C}_{n,M}$ of $(\mathbb{U}^-, \mathbb{U}^+)$.
3. Replace g by the uncostly monotonic (increasing) function
$$\tilde{g}(\mathbf{x}) \quad = \quad \begin{cases} -1 & \text{if } P(g(\mathbf{x}) \leqslant 0) | \hat{C}_{n,M}) \geqslant 1/2, \\ +1 & \text{else.} \end{cases}$$
4. Sample $\mathbf{x_1}, \ldots, \mathbf{x_Q}\mathbf{iid}\sim\mathcal{U}_\mathbb{U}$ and compute $\tilde{p}_{n,M,Q} = Q^{-1} \sum_{k=1}^{Q} \mathbb{1}_{\{\tilde{g}(\mathbf{x_k}) \leqslant 0\}}$.
5. For $i = 1, \ldots, S$, get a MLE estimator $\tilde{p}_{n,M,Q}^{(i)}$ then estimate B_n by
$$\hat{B}_{n,M,Q,S} \quad = \quad S^{-1} \sum_{i=1}^{S} \tilde{p}_{n,M,Q}^{(i)} - \tilde{p}_{n,M,Q}.$$

Numerical tests in function of n and d were conducted, and the results are presented in Table 4.2. The bias correction is found to be effective even from a moderate number of iterations (some hundreds) until dimension 5, and a budget of at least $n = 1,000$ is enough to correct a bias in dimension 8. With less than 10% of overestimation on average on this example, these bootstrap heuristics also appear relevant when the exact value of p is less interesting than its magnitude, which is often the case in design optimization where it is aimed to diminish p of a given factor by constraining the inputs [38].

Table 1. — Relative error in % between estimated bias and real bias, for two true probabilities $p = 5\%$ and $p = 0.5\%$. Results are averaged on 100 experiments, each boostrap estimation being based on $S = 1,000$ MLE replicates. For each n, the neural network is build from $M = 10^6$ sampled vectors, with a classification error rate less than 0.25% on these training data.

	Dimension d								
	$p = 0.05$					$p = 0.005$			
n	2	3	4	5	8	2	3	4	5
50	1.19	4.17	6.93	16.87	27.85	8.21	13.80	18.22	39.57
100	0.28	2.31	4.79	12.94	22.98	6.15	11.55	15.67	31.40
250	0.21	1.87	3.34	8.74	19.12	3.28	8.72	11.01	24.68
500	0.12	1.25	2.87	6.20	16.76	1.14	5.84	8.12	16.06
1000	-0.02	0.47	2.14	2.97	12.85	0.12	2.72	5.28	9.23
2000	-0.008	-0.28	1.61	2.08	7.66	-0.34	1.55	3.09	6.51

5. Numerical experiments II: a simplified hydraulic case-study

Several authors in structural reliability [13, 26, 32] considered a simplified but realistic hydraulic model linking the downstream water level H (m) of a river section, of width $b = 300$ (m) and length $l = 5000$ (m), with the upstream discharge Q (m^3/s) and the friction coefficient K_s $(m^{1/3}/s)$ of the river bed. Denoting Z_m and Z_v the upstream and downstream altitude of the river bed above seal level,

$$H \quad = \quad \left(\frac{Q}{bK_s \sqrt{\frac{Z_m - Z_v}{l}}} \right)^{3/5}.$$

Assuming a dike level $h_0 = 55.5$ (m), the flood probability is $p = P(g' \circ F^{-1}(\mathbf{X}) \leqslant 0)$ where $\mathbf{Y} = \{F_Q(Q), K_s\}$ (2-dim. version) or $\mathbf{Y} = \{Q, K_s, Z_m, Z_v\}$ (4-dim. version), F is the cdf of $\mathbf{Y}$ and

$$g'(\mathbf{Y}) \quad = \quad h_0 - Z_v - H(\mathbf{Y}),$$

which is increasing in $(-Q, K_s, Z_m, -Z_v)$. Input distributions or point values are chosen as in [26]. Q follows a Gumbel distribution with location 1013 and scale 558, truncated in $[10, 10^4]$. K_s is normal $\mathcal{N}(27.8, 3^2)$ truncated in 0. In the 4-dim. version, Z_m and Z_v are triangular on $[53.5, 56.5]$ and

$[48.5, 51.5]$ with respective modes 55 and 50 (their respective point values in the 2-dim. version).

For several computational budgets, and averaged over 100 repeated experiments, two alternative methods are compared to the MRM bias-corrected MLE: the MC method and an engineering FORM-IS method build on two steps: (a) with a limited number of trials (no more than 40), the First-Order Reliability Method (FORM) is run to provide an estimate of the conception point $\beta = \arg\min \|\mathbf{u}\|$ on $\{g' \circ T^{-1} \circ \Phi(\mathbf{u}) \leqslant 0\}$, with Φ the standard normal pdf and $\mathbf{u}$ a random variable evolving in the $d-$dimensional standard Gaussian space U ; (b) an Importance Sampling (IS) method that uses the budget left to sample in U using a standard normal distribution centered on β. See [1] for details about the implementation of the method.

For a given n, the three methods are compared through the following indicators: $\mathrm{E}[\hat{p}_n]$, $\mathrm{CV}[\hat{p}_n]$ and the relative average precision $\gamma_n = \mathrm{E}[(p_n^+ - p_n^-)]/p$. Using the DOEs produced by the MC and the FORM-IS methods, these bounds can obviously be computed accounting for the monotonicity of g. For each version a MC computation involving 40,000 particules provides a precise estimate of p, which is used for estimating p in γ_n. Finally, $S = 1,000$ bootstrap replicates were used for the correction of each MRM-MLE estimate. The results are summarized on Table 2.

Table 2. — Estimation results for the two-dimensional and four-dimensional versions of the problem.

n	method	dimension = 2			dimension = 4		
		$\mathrm{E}[\hat{p}_n]$	$\mathrm{CV}[\hat{p}_n]$	γ_n	$\mathrm{E}[\hat{p}_n]$	$\mathrm{CV}[\hat{p}_n]$	γ_n
100	MC	0.002775	190%	2,900%	0.010075	99%	
	FORM-IS	0.002241	68%	478%	0.018147	74%	
	MRM	0.002781	14%	48%	0.015498	82%	1,400%
200	MC	0.002775	134%	630%	0.010075	70%	2,300%
	FORM-IS	0.002667	44%	244%	0.010242	42%	2,230%
	MRM	0.002776	6%	24%	0.012451	35%	800%
1,000	MC	0.002775	60%	515%	0.010075	31%	1,200%
	FORM-IS	0.002736	27%	168%	0.009959	27%	1,000%
	MRM	0.002775	0.12%	5.6%	0.010911	20%	300%
40,000	MC	0.002775	9.5%	475%	0.010075	5%	247%

In terms of magnitude, the three methods perform similarly. The benefit of using MRM instead of MC or FORM-IS in these low dimensions clearly appears in most cases, and more obviously in dimension 2: MC needs at least 200 times more iterations than MRM to reach a similar precision $\mathrm{CV}[\hat{p}_n]$, and if FORM-IS is significantly better than MC, the precision of its estimates remains far beyond of those produced by MRM. In dimension 4, the difference between these two methods somewhat vanishes and they lead to close performances when the number of iterations remains low. For both dimensional cases, it was noticed that a single FORM run can provide a crude estimate of p with good magnitude after 10 iterations only. But the dimensional increasing allows the part of the importance sampling falling into the non-dominated area to be greater than in a two-dimensional setting.

6. Discussion

Many structural reliability problems deal with the fast estimation of a probability p of an undesirable event. This event can often be defined by the occurence of an exceedance in output of some time-consuming function g with stochastic multidimensional inputs. In the present article, g is assumed to be monotonic and possibly non-continuous in a non-empty set.

Pursuing pioneering works by de Rocquigny [13] and Limbourg *et al.* [26] who explored heuristically the benefits of this framework, this article first offers a formal description of the so-called Monotonic Reliability Method (MRM) that focuses on the existence of deterministic bounds around p. A sequential strategy of numerical experiments in the input space allows for a progressive narrowing of this interval. The second and main aspect of the paper is the definition and the study of a statistical estimator of p when the strategy becomes stochastic and leans on uniform nested sampling. Easy to compute, it is defined as the maximizer of a likelihood (MLE) of dependent data sampled from Bernoulli distributions, whose parameters are explicit functions of the dynamic bounds.

A keypoint of the paper is the theoretical description of its asymptotic properties, which are found similar to those arising from the classical estimation theory, provided some intuitive assumptions are respected. They are found mild in practice on some examples. Both theoretical and applied results show a significant improvement of the fastness and the robustness of this estimator with respect to the usual Monte Carlo estimator. In the third part of the paper, boostrap heuristics are proposed and carried out successfully to remove the non-asymptotic bias affecting the MLE, via constrained neural networks. Only a basic continuity assumption on the limit state (or *failure*) surface is needed to benefit from their universal approximation capabilities.

Thus, the tools proposed in this article and its supplementary material in Appendix can be directly used in structural reliability applications, without preliminary learning step (as usual, for instance, in stratification methods). However, the generality of the frame allows for a wider range of theoretical and applied studies. These research avenues are briefly discussed in the following items.

Bias correction. — Following [21], support vector machines (SVM) should probably be considered instead of neural networks, since their geometric interpretation of margin maximizers appears more relevant to the problem than neural networks. In addition to the monotonicity constraint, they should be build at step n under the linear constraint that the volume under the predicted surface be equal to the current (biased) estimator $\hat{p}_n$. This would certainly improve the properties of the bootstrap heuristics. More importantly, this method should be now tested on a large variety of examples, and the intuitive feeling of its ability to correct the bias must be confirmed by more applied and theoretical studies.

In parallel, future studies should focus on adopting a corrective approach to the bias affecting the MLE, then on selecting a slippery window of indexes, according to (4.1) or a similar rule, such that the MLE converges faster to p. The comparison of the experimental benefits of both approaches would help the method to become more ready-to-use.

Simplifying the assumptions. — Most of the technical assumptions that are needed to get the theoretical results present some intuitive features, and are underlyingly linked to the nature of the limit state surface. However, they remain difficult to check in practice, although asymptotic normality was always noticed in numerical experiments. Therefore, future work should be dedicated to simplifying those assumptions and classifying the limit state surfaces in function of their ability to allow a fast and robust estimation of p.

Sensitivity studies. — Crucial tasks in structural reliability are sensitivity studies of probabilistic indicators to the uncertainty input model [31]. Therefore, assuming $\mathbf{X} = F(\mathbf{Y})$ where the $\mathbf{Y}$ represent *physical* inputs with cdf F, given a budget n, the variations of $(p_n^-, p_n^+, \hat{p}_n)$ due to modifying F in F_ϵ should be the subject of future works. As a supplementary benefit of the method, the new values $(p_{k,\epsilon}^-, p_{k,\epsilon}^+, \hat{p}_{n,\epsilon})$, for $k \in \{0, \dots, n\}$, can be recomputed without any supplementary call to g, thanks to an importance sampling mechanism. Indeed, as the subspaces $(\mathbb{U}_k^-, \mathbb{U}_k^+)$ remain dominated whatever the choice made on input distributions in the physical space, then

$$\begin{aligned} p^{-}_{k,\epsilon} &= \int_{F^{-1}(\mathbb{U}^{-}_{k})} dF_{\epsilon}(\mathbf{y}) \\ \text{and} \quad p^{+}_{k\epsilon} &= 1 - \int_{F^{-1}(\mathbb{U}^{+}_{k})} dF_{\epsilon}(\mathbf{y}), \end{aligned}$$

which can computed by a simple Monte Carlo method. In such future studies, we suggest that the progressive bounds could be defined as *robust* if they remain true whatever the fluctuations of F_ϵ in a well-funded variational class around F.

Exploring other forms of stochastic DOEs. — A keypoint of future works will be to elaborate unbiased estimators from sequential stochastic designs of experiments with non-asymptotic properties. Indeed, the asymptotic variance of the MLE reaches the Cramer-Rao bound $J_n^{-1}(p)$. Therefore any unbiased estimator based on sequential uniform sampling, especially those defined by $\tilde{p}_n = \sum_{k=1}^{n} \omega_k p_k$ where the ω_k are now *deterministic* weights, independent on p and summing to 1, will never reach a lower variance than $J_n^{-1}(p)$, even though the ω_k are optimized to this aim. Improving the Monte Carlo acceleration $nJ_n^{-1}(p)/(p(1-p))$ will only be possible using less naive strategies than uniform samplings. The problem of defining such samplings so that an unbiased estimator of p has better statistical properties will be the subject of a future article.

Towards partial monotonicity. — Finally, the practical limits of monotonicity assumptions should be refined. Intuitively, monotonicity as a building hypothesis seems antagonist to high-dimensional structural safety problems, and could mainly characterize the behavior of g as a function of its most prominent input variables (as it could be measured through global sensitivity analysis). Indeed, the real examples treated in [13, 26] and [34] do not go beyond dimension 4. Partial monotonicity, as defined in [11], seems a more appealing and realistic property, for which the methods developed in a pure monotonicity context should be adapted in the future.

7. Appendix A: proofs

Proof of Lemma 3.1. — An infinite uniform sampling on $\mathbb{U}$ provides on the open sets $(\mathbb{U}^-, \mathbb{U}^+)$ two topologies constituted by the collections of open subsets $\mathbb{U}_0^-, \ldots, \mathbb{U}_n^-, \ldots$, and $\mathbb{U}_0^+, \ldots, \mathbb{U}_n^+, \ldots$. Hence the sequence $(\mathbb{U}_n^-, \mathbb{U}_n^+)$ define two covers (exhaustions) of $(\mathbb{U}^-, \mathbb{U}^+)$. Then

$$U_\infty^- = \bigcup_{k=0}^{\infty} U_k^- \;=\; \mathbb{U}^-, \qquad U_\infty^+ = \bigcup_{k=0}^{\infty} U_k^+ \;=\; \mathbb{U}^+$$

and $\lim_{n\to\infty} p_n^- = \lim_{n\to\infty} P(\mathbf{X} \in \mathbb{U}_n^-) = P(\mathbf{X} \in U_\infty^-) = p$ by inclusion. Similarly, $\lim p_n^+ = p$. Furthermore, given p_0^- and p_0^+, p_n^- and $1 - p_n^+$ are $\mathcal{F}_{n-1}-$adapted submartingales bounded in $\mathbb{L}_p$ $\forall p \geqslant 1$. Then, from Doob's theorem [30], the bounds converge almost surely to p. □

Proof of Proposition 3.2. — One may write $\ell_n''(p) = \sum_{k=1}^n \tilde{\omega}_k(p)\, S_k(p)$ with

$$\begin{aligned} S_k(p) &= -1 + (p_k - p)\,\tilde{\omega}_k(p)\left(2p - p_{k-1}^- - p_{k-1}^+\right), \qquad (7.1)\\ &= -\tilde{\omega}_k(p)(p - p_k)^2. \end{aligned}$$

Hence $\ell_n''(p) < 0$ in (p_{n-1}^-, p_{n-1}^+). Besides, $\lim_{p\to p_{n-1}^-} \ell_n'(p) = \infty$ and $\lim_{p\to p_{n-1}^+} \ell_n'(p) = -\infty$. Hence, by twice continuity and differentiability of $\ell_n(p)$, the mean value theorem implies the existence and unicity of a MLE $\hat{p}_n$ in $]p_{n-1}^-, p_{n-1}^+[$. □

Proof of Lemma 3.3. — We shall proceed by induction. Since $p_0^- < p < p_0^+$, (3.5) and (3.4) hold for $n = 0$. Denote $\eta_n = 1/(p - p_n^-)^2$. For $n \geqslant 1$, it is assumed that $\mathrm{E}[\eta_n] < \infty$. Then

$$\mathrm{E}\left[\eta_{n+1}\right] = \mathrm{E}\left[\eta_n \mathrm{E}\left[1 - \xi_{\mathbf{x_{n+1}}} | \mathcal{F}_n\right]\right] + \mathrm{E}\left[\mathrm{E}\left[\xi_{\mathbf{x_{n+1}}} / \left(p - p_n^- - \mathrm{Vol}_{\mathbf{x_{n+1}}}^-\right)^2 | \mathcal{F}_n\right]\right]$$

with $\mathrm{Vol}_{\mathbf{x_{n+1}}}^- = \int_{\mathbb{U}_n \cap \mathbb{U}^-} \mathbb{1}_{\{\mathbf{x} \preceq \mathbf{x_{n+1}}\}}\, d\mathbf{x}$ the additive volume of formerly non-dominated failure points in $\mathbb{U}_n$ that are now dominated by the failure point $\mathbf{x_{n+1}}$. By hypothesis, the first term is always finite. Furthermore, with

$$\mathrm{Vol}_{\mathbf{x_{n+1}}}^- \leqslant \sup_{\mathbf{x_n} \in \mathbb{U}_n \cap \mathbb{U}^-} \int_{\mathbb{U}_n \cap \mathbb{U}^-} \mathbb{1}_{\{\mathbf{x} \preceq \mathbf{x_n}\}}\, d\mathbf{x} \;=\; \sup_{\mathbf{x_n} \in \mathcal{S}} \int_{\mathbb{U}_n \cap \mathbb{U}^-} \mathbb{1}_{\{\mathbf{x} \preceq \mathbf{x_n}\}}\, d\mathbf{x},$$

(2.2) implies that $\mathrm{Vol}_{\mathbf{x_{n+1}}}^- < p - p_n^-$. Since $p_n^- = p_{n-1}^- + \mathrm{Vol}_{\mathbf{x_n}}^-$, etc., one has $\sum_{k=1}^{n+1} \mathrm{Vol}_{\mathbf{x_k}}^- < p - p_0^-$. Then

$$\mathrm{E}\left[\mathrm{E}\left[\xi_{\mathbf{x_{n+1}}}/\left(p-p_n^- - \mathrm{Vol}^-_{\mathbf{x_{n+1}}}\right)^2|\mathcal{F}_n\right]\right] \;=\; \mathrm{E}\left[\xi_{\mathbf{x_{n+1}}}/\left(p-p_0^- - \sum_{k=1}^{n+1}\mathrm{Vol}^-_{\mathbf{x_k}}\right)^2\right] \;<\; \infty.$$

The same rationale applies to $1/(p_n^+ - p)^2$, by symmetry, since $p_{n+1}^+ = p_n^+ - \mathrm{Vol}^+_{\mathbf{x_{n+1}}}$ with $\mathrm{Vol}^-_{\mathbf{x_{n+1}}} = \int_{\mathbb{U}_n\cap\mathbb{U}^+} \mathbb{1}_{\{1-\mathbf{x}\preceq 1-\mathbf{x_{n+1}}\}}\, d\mathbf{x}$. □

Proof of Proposition 3.4. — One has $\mathrm{E}[\ell_n'(p)] = \sum_{k=1}^n \mathrm{E}\left[\tilde{\omega}_k(p)\mathrm{E}\left[p_k - p|\mathcal{F}_{k-1}\right]\right] = 0$ since $\tilde{\omega}_{n+1}$ depends only on $\mathcal{F}_n$, hence the Fisher information $J_n(p) = \mathrm{Var}[\ell_n'^2(p)] = \mathrm{E}[\ell_n'^2(p)]$ is equal to $-\mathrm{E}[\ell_n''(p)]$ by twice differentiability and continuity of $\ell_n(\cdot)$, similarly to a classic iid. case. Assumption 4 implies that $\forall n < \infty$, $p_{n-1}^- < p < p_{n-1}^+$, ie. p cannot be reached in any finite number of iterations, so that these quantities are well defined. With $-S_n(p) = \tilde{\omega}_n(p)(p-p_n)^2 \;\forall\; n \geqslant 0$ from (7.1),

$$J_n(p) \;=\; \sum_{k=1}^n \mathrm{E}\left[\tilde{\omega}_k^2(p)\mathrm{Var}\left[p_k|\mathcal{F}_{k-1}\right]\right] \;=\; \sum_{k=1}^n \mathrm{E}\left[\tilde{\omega}_k(p)\right]$$

since

$$\begin{aligned}\mathrm{Var}\left[p_n|\mathcal{F}_{n-1}\right] &= \left(p_{n-1}^+ - p_{n-1}^-\right)^2 \mathrm{E}\left[\xi_{\mathbf{x_n}}|\mathcal{F}_{n-1}\right] - \left(p-p_{n-1}^-\right)^2,\\ &= \left(p_{n-1}^+ - p_{n-1}^-\right)\left(p-p_{n-1}^-\right) - \left(p-p_{n-1}^-\right)^2,\\ &= \tilde{\omega}_n^{-1}(p). \qquad (7.2)\end{aligned}$$

Inequality (3.6) is a simple consequence of Jensen's inequality:
since $\mathrm{E}^{-1}\left[\tilde{\omega}_k^{-1}(p)\right] \leqslant \mathrm{E}\left[\tilde{\omega}_k(p)\right]$,

$$\text{then } J_n^{-1}(p) \;\leqslant\; \left(\sum_{k=1}^n \mathrm{E}^{-1}\left[\tilde{\omega}_k^{-1}(p)\right]\right)^{-1} \;=\; \frac{p(1-p)}{\sum_{k=1}^n (1-c_{k-1})^{-1}}. \qquad \square$$

Proof of Proposition 3.5. — Using the notation $S_k(p)$ defined in (7.1),

$$J_n(p) \;=\; -\mathrm{E}\left[\frac{p(1-p)}{nV_n^{MC}(p)}\sum_{k=1}^n \tilde{\omega}_k\left(p\right) S_k\left(p\right)\right] \;=\; n^{-1}\frac{\tilde{J}_n(p)}{V_n^{MC}(p)}$$

with

$$\tilde{J}_n(p) \;=\; \mathrm{E}\left[\sum_{k=1}^n p(1-p)\left(p_{k-1}^+ - p\right)^{2\xi_{\mathbf{x_k}}-2}\left(p-p_{k-1}^-\right)^{-2\xi_{\mathbf{x_k}}}\right],$$

which can be rewritten as

$$\tilde{J}_n(p) = \sum_{k=1}^{n} \left\{ \mathrm{E}\left[\xi_{\mathbf{x_k}} \left(\frac{p(1-p)}{\left(p - p_{k-1}^-\right)^2} \right) \right] + \mathrm{E}\left[(1-\xi_{\mathbf{x_k}}) \left(\frac{p(1-p)}{\left(p_{k-1}^+ - p\right)^2} \right) \right] \right\}.$$

Since $p^{-1}\left(\left(p_{k-1}^+ - p\right) + \left(p - p_{k-1}^-\right)\right) \leqslant \rho_{k-1}$, then

$$\begin{aligned} p_{k-1}^+ - p &\leqslant p(\rho_{k-1} - 1) + p_{k-1}^- \leqslant p(\rho_{k-1} - 1) + p = p\rho_{k-1}, \\ p - p_{k-1}^- &\leqslant p(\rho_{k-1} + 1) - p_{k-1}^+ \leqslant p(\rho_{k-1} + 1) - p = p\rho_{k-1}. \end{aligned}$$

Hence

$$\frac{p(1-p)}{\left(p_{k-1}^+ - p\right)^2} \geqslant \frac{1-p}{p\rho_{k-1}^2} \quad \text{and} \quad \frac{p(1-p)}{\left(p - p_{k-1}^-\right)^2} \geqslant \frac{1-p}{p\rho_{k-1}^2}.$$

Consequently,

$$\begin{aligned} \tilde{J}_n(p) &\geqslant \frac{1-p}{p} \sum_{k=1}^{n} \mathrm{E}\left[\frac{1}{\rho_{k-1}^2} \left(\xi_{\mathbf{x_k}} + 1 - \xi_{\mathbf{x_k}} \right) \right], \\ &\geqslant \frac{1-p}{p} n \mathrm{E}\left[\frac{1}{\rho_0^2} \right] \end{aligned}$$

since (ρ_n) is a strictly decreasing positive process. Since (p_n^-, p_n^+) are predictible processes, p_0^- and p_0^+ are deterministic quantities, then

$$\mathrm{E}\left[\frac{1}{\rho_0^2} \right] = \left(\frac{p_0^-}{p_0^+ - p_0^-} \right)^2 = 1/\gamma_0,$$

and $\tilde{J}_n(p) \geqslant n\left(\frac{1-p}{\gamma_0 p}\right)$ which proves (3.7). □

Proof of Theorem 3.6. — Given the strong consistency of $\hat{p}_n$, its asymptotic normality can be established using arguments studied in [9, 10]. Showing that $\ell_n'(p)$ is a $\mathcal{F}_{n-1}-$adapted martingale is a classic result:

$$\mathrm{E}\left[\ell_{n+1}'(p) - \ell_n'(p) | \mathcal{F}_n \right] = \tilde{\omega}_{n+1}(p) \mathrm{E}\left[p_{n+1} - p | \mathcal{F}_n \right] = 0.$$

Furthermore $J_n(p) < n\mathrm{E}[\tilde{\omega}_n(p)] < \infty$ under Assumption 4 (cf. Lemma 3.3) Hence $\ell_n'(p)$ is square integrable. Denoting $\Delta_n(p) = \ell_n'(p) - \ell_{n-1}'(p)$, then $\Delta_n^2(p) = \tilde{\omega}_n^2(p)(p_n - p)^2$ and

$$\mathrm{E}\left[\Delta_n^2(p) | \mathcal{F}_{n-1} \right] = \tilde{\omega}_n^2(p) \mathrm{Var}\left[p_n | \mathcal{F}_{n-1} \right] = \tilde{\omega}_n(p).$$

Then $< \ell'(p) >_n = \sum_{k=1}^n \tilde{\omega}_k(p)$ denotes the increasing (or bracket) process of $\ell'_n(p)$. The proof can be achieved in three steps.

1) With $J_n(p) = \mathrm{E}[< \ell'(p) >_n]$ and $\lim_{n\to\infty} J_n(p) = \infty$ from (7.7), establishing asymptotic normality first requires to prove the following law of large numbers (LLN)

$$M_n(p) \;=\; J_n^{-1}(p) < \ell'(p) >_n -1 \quad \xrightarrow{I\!P} \quad 0. \tag{7.3}$$

Denote $W_n(p) = \sum_{k=1}^n (\tilde{\omega}_k(p) - \mathrm{E}[\tilde{\omega}_k(p)])$ Then, $\forall \epsilon > 0$,

$$\begin{aligned} P\left(|M_n(p)| > \epsilon\right) &= P\left(J_n^{-1}(p)|W_n(p)| > \epsilon\right), \\ &\leqslant P\left(V_n^{MC}(p)\frac{p\gamma_0}{1-p}|W_n(p)| > \epsilon\right) \quad \text{from (3.7)}, \\ &\leqslant P\left(\frac{1}{n}|W_n(p)| > \epsilon'\right) \quad \text{with } \epsilon' = \epsilon/(p^2\gamma_0), \end{aligned}$$

which tends to 0 under (i) and proves (7.3).

2) For all $k \in \{1, \ldots, n\}$, denote $\Gamma_{k,n} = J_n^{-1/2}(p)|\Delta_k(p)|$. The second requirement of asymptotic normality is proving the following Lindeberg condition: $\forall \epsilon > 0$,

$$\frac{1}{J_n(p)} \sum_{k=1}^n \mathrm{E}\left[\Delta_k^2(p) \mathbb{1}_{\{\Gamma_{k,n} > \epsilon\}} | \mathcal{F}_{k-1}\right] \quad \xrightarrow[n\to\infty]{I\!P} \quad 0. \tag{7.4}$$

A Lyapunov condition is often used instead of (7.4), but requires $2+\delta$-order moment assumptions on $\tilde{\omega}_k(p)$. An alternative approach is the following. From Markov's inequality and since the $\tilde{\omega}_k(p)$ are increasing functions of k,

$$P\left(\Gamma_{k,n} > \epsilon | \mathcal{F}_{k-1}\right) \;\leqslant\; \frac{\tilde{\omega}_k(p)}{\epsilon^2 J_n(p)} \;\leqslant\; \frac{\tilde{\omega}_n(p)}{\epsilon^2 J_n(p)}.$$

It follows from (7.3) that

$$\begin{aligned} \frac{\tilde{\omega}_n(p)}{J_n(p)} + \frac{1}{J_n(p)} \sum_{k=1}^{n-1} \tilde{\omega}_n(p) \quad &= \quad \frac{\tilde{\omega}_n(p)}{J_n(p)} + \left(\frac{J_{n-1}(p)}{J_n(p)}\right)\left(\frac{< \ell'(p) >_{n-1}}{J_{n-1}(p)}\right), \\ &\xrightarrow[n\to\infty]{I\!P} \quad 1. \end{aligned}$$

However, by Lemma 3.3, $\mathrm{E}[\tilde{\omega}_n(p)] < \infty$ which means that $J_n(p) \overset{\infty}{\sim} J_{n-1}(p)$. Necessarily, $\tilde{\omega}_n(p)/J_n(p) \xrightarrow[n\to\infty]{I\!P} 0$ and

$$L_{k,n} \;=\; \mathrm{E}\left[\mathbb{1}_{\{\Gamma_{k,n} > \epsilon\}} | \mathcal{F}_{k-1}\right] \quad \xrightarrow[n\to\infty]{I\!P} \quad 0. \tag{7.5}$$

Note besides that

$$\begin{aligned} & \mathrm{E}\left[\Delta_k^2(p)\mathbb{1}_{\{\Gamma_{k,n}>\epsilon\}}|\mathcal{F}_{k-1}\right] \\ \leqslant\ & \mathrm{E}\left[\Delta_k^2(p)|\mathcal{F}_{k-1}\right]\mathrm{E}\left[\mathbb{1}_{\{\Gamma_{k,n}>\epsilon\}}|\mathcal{F}_{k-1}\right] + \left|\mathrm{Cov}\left[\Delta_k^2(p),\mathbb{1}_{\{\Gamma_{k,n}>\epsilon\}}|\mathcal{F}_{k-1}\right]\right|, \\ \leqslant\ & \tilde{\omega}_k(p)L_{k,n} + \tilde{\omega}_k^2(p)\sqrt{\mathrm{Var}\left[(p_k-p)^2|\mathcal{F}_{k-1}\right]}\sqrt{\mathrm{Var}\left[\mathbb{1}_{\{\Gamma_{k,n}>\epsilon\}}|\mathcal{F}_{k-1}\right]} \end{aligned}$$

from Cauchy-Schwarz inequality. Since $\mathrm{Var}[X^2] \leqslant \mathrm{E}[X]$ when $X \in \{0,1\}$, then $\sqrt{\mathrm{Var}[\mathbb{1}_{\{\Gamma_{k,n}>\epsilon\}}|\mathcal{F}_{k-1}]} \leqslant \sqrt{L_{k,n}}$. Furthermore, denote

$$K_{k,n}(p) \quad = \quad \tilde{\omega}_k(p)\sqrt{\mathrm{Var}\left[(p_k-p)^2|\mathcal{F}_{k-1}\right]}.$$

From Lemma 7.1 below and under (ii), then $K_{k,n}(p) \xrightarrow{I\!P} 0$. Therefore, one may write

$$\mathrm{E}\left[\Delta_k^2(p)\mathbb{1}_{\{\Gamma_{k,n}>\epsilon\}}|\mathcal{F}_{k-1}\right] \quad \leqslant \quad \tilde{\omega}_k(p)\beta_{k,n}$$

with $\beta_{k,n} = L_{k,n} + K_{k,n}(p)\sqrt{L_{k,n}} \xrightarrow{I\!P} 0$ from (7.5). Then

$$\frac{1}{J_n(p)}\sum_{k=1}^{n}\mathrm{E}\left[\Delta_k^2(p)\mathbb{1}_{\{\Gamma_{k,n}>\epsilon\}}|\mathcal{F}_{k-1}\right] \leqslant \frac{<\ell'(p)>_n}{J_n(p)}\frac{\sum_{k=1}^{n}\tilde{\omega}_k(p)\beta_{k,n}}{\sum_{k=1}^{n}\tilde{\omega}_k(p)} \tag{7.6}$$

and given (7.3), Toeplitz lemma proves (7.4). Finally, (7.3) and (7.4) prove the two martingale central limit theorems [2]:

$$J_n^{-1/2}(p)\ell_n'(p) \quad \xrightarrow[n\to\infty]{\mathcal{L}} \quad \mathcal{N}(0,1), \tag{7.7}$$

$$\frac{\sqrt{J_n(p)}}{<\ell'(p)>_n}\ell_n'(p) \quad \xrightarrow[n\to\infty]{\mathcal{L}} \quad \mathcal{N}(0,1). \tag{7.8}$$

Lemma 7.1. — *If* $\exists\ \gamma_\infty$ *such that* $0 < \gamma_\infty < \infty$ *and* $\frac{p_n^+ - p}{p - p_n^-} \xrightarrow{I\!P} \gamma_\infty$, *then,* $\forall k \geqslant 1$,

$$Var\left[\tilde{\omega}_k(p)(p_k-p)^2|\mathcal{F}_{k-1}\right] \quad \xrightarrow{I\!P} \quad \gamma_\infty + 1/\gamma_\infty - 2.$$

Proof.— One may write

$$\begin{aligned} \tilde{\omega}_k(p)(p_k-p)^2 \quad &= \quad (1-\xi_{\mathbf{x_k}})\frac{p-p_{k-1}^-}{p_{k-1}^+-p} + \xi_{\mathbf{x_k}}\frac{p_{k-1}^+-p}{p-p_{k-1}^-}, \\ &= \quad \xi_{\mathbf{x_k}}\left[\tilde{\omega}_k(p)\left\{\left(p_{k-1}^+-p\right)^2 - \left(p-p_{k-1}^-\right)^2\right\}\right] + \frac{p-p_{k-1}^-}{p_{k-1}^+-p} \end{aligned}$$

With $\xi_{\mathbf{x_k}} \sim \mathcal{B}(\gamma_k)$ *and from (3.1), then*

$$\begin{aligned} &Var\left[\tilde{\omega}_k(p)(p_k - p)^2 | \mathcal{F}_{k-1}\right] \\ &\quad = \frac{\tilde{\omega}_k(p)}{(p^+_{k-1} - p^-_{k-1})^2}\left[(p^+_{k-1} + p^-_{k-1} - 2p)(p^+_{k-1} - p^-_{k-1})\right]^2, \\ &\quad = \tilde{\omega}_k(p)(p^+_{k-1} + p^-_{k-1} - 2p)^2, \\ &\quad = \frac{p^+_{k-1} - p}{p - p^-_{k-1}} + \frac{p - p^-_{k-1}}{p^+_{k-1} - p} - 2 \xrightarrow{I\!P} \gamma_\infty + 1/\gamma_\infty - 2. \end{aligned}$$

3) A last condition is required to transfer the asymptotic normality from $\ell'_n(p)$ to $(\hat{p}_n - p)$. Since $p^-_{n-1} < \hat{p}_n < p^+_{n-1}$, for any n there always exists an open neighborhood $\mathcal{V}_{\hat{p}_n}$ of p containing $\hat{p}_n$. From twice differentiability of $\ell_n(\cdot)$ and continuity of $\ell'_n(\cdot)$ in $\mathcal{V}_{\hat{p}_n}$, the mean value theorem implies there exists some intermediate point $\bar{p}_n \in \mathcal{V}_{\hat{p}_n}$ between p and $\hat{p}_n$ such that

$$\ell'_n(\hat{p}_n) \;=\; 0 \;=\; \ell'_n(p) + (\hat{p}_n - p)\,\ell''_n(\bar{p}_n)$$

and moreover $\bar{p}_n \xrightarrow{a.s.} p$. Thus, with $\ell''_n(\tilde{p}_n) \neq 0$,

$$(\hat{p}_n - p) \;=\; \ell'_n(p)\left(-\ell''_n(\bar{p}_n)\right)^{-1} \tag{7.9}$$

and it is necessary to prove the LLN

$$\frac{\ell''_n(\bar{p}_n)}{< \ell'(p) >_n} \;\xrightarrow{I\!P}\; 1 \tag{7.10}$$

to obtain the final result (Theorem 3 in [10]), combining (7.10) with (7.3) and (7.7). Based on (iii) this last LLN is straightforward. Indeed, $\forall k \leqslant n$,

$$\left|\frac{p^+_k - \bar{p}_n}{p^+_k - p} - 1\right| \;=\; \frac{|\bar{p}_n - p|}{p^+_k - p} \;\leqslant\; \frac{|\bar{p}_n - p|}{p^+_n - p} \xrightarrow{I\!P} 0.$$

Similarly $(\bar{p}_n - p^-_k)/(p - p^-_k) - 1 \underset{k\to n\to\infty}{\xrightarrow{I\!P}} 0$. With $p_{k+1} \in \{p^-_k, p^+_k\}$ then, for $k \in \{0, \ldots, n-1\}$,

$$\gamma_{k+1,n} \;=\; (\bar{p}_n - p_{k+1})^2\,\tilde{\omega}_{k+1}(\bar{p}_n) \underset{k\to n\to\infty}{\xrightarrow{I\!P}} 1.$$

Furthermore, some calculus proves that $\kappa_{k,n} = \tilde{\omega}_{k+1}(\bar{p}_n)/\tilde{\omega}_{k+1}(p) \underset{k\to n\to\infty}{\xrightarrow{I\!P}} 1$. Then, with $-\ell''_n(p) = \sum_{k=1}^n \Delta^2_k(p)$,

$$\frac{\ell''_n(\bar{p}_n)}{< \ell'(p) >_n} \;=\; \frac{\sum_{k=1}^n \tilde{\omega}_k(p)\kappa_{k,n}\gamma_{k,n}}{\sum_{k=1}^n \tilde{\omega}_k(p)} \underset{k\to n\to\infty}{\xrightarrow{I\!P}} 1 \quad \text{from Toeplitz lemma.}$$

□

Proof of Proposition 3.7. — Using the notations of the previous proof, note that $\hat{J}_n(p) =< \ell'(p) >_n$. By twice continuity and derivability of $\hat{J}_n^{-1}(.)$ in $]p_n^-, p_n^+[$, a Taylor expansion gives

$$\hat{J}_n^{-1}(\hat{p}_n) \quad = \quad \hat{J}_n^{-1}(p) - \frac{\hat{J}_n'(p)}{\hat{J}_n^2(p)} (\hat{p}_n - p)(1 + o(1)).$$

After some calculus,

$$\frac{\hat{J}_n^{5/2}(p)}{|\hat{J}_n'(p)|} \left(\hat{J}_n^{-1}(\hat{p}_n) - J_n^{-1}(p) \right) \quad = \quad R_n U_n + R_n Z_n$$

with $R_n = \sqrt{J_n(p)/\hat{J}_n(p)} \xrightarrow{I\!P} 1$ from (7.3), $U_n = \mathrm{sgn}(\hat{J}_n'(p))\sqrt{J_n(p)}\,(\hat{p}_n - p)\,(1+o(1)) \xrightarrow{\mathcal{L}} \mathcal{N}(0,1)$ from Theorem 3.6, and

$$Z_n \quad = \quad \frac{\hat{J}_n^{3/2}(p)}{|\hat{J}_n'(p)|} \left(\frac{\hat{J}_n(p)}{J_n(p)} - 1 \right).$$

Thanks to Slutsky's theorem, it is enough to show that $Z_n \xrightarrow{I\!P} 0$ to prove the statement of the proposition. Notice that

$$\hat{J}_n'(p) \quad = \quad \sum_{k=1}^{n} \tilde{\omega}_k^2(p) \left\{ 2p - \left(p_{k-1}^+ + p_{k-1}^- \right) \right\}$$

which is always nonzero assuming (iv). Hölder's inequality gives

$$\frac{\sum_{k=1}^{n} \tilde{\omega}_k(p)}{\sum \tilde{\omega}_k^2(p) \left\{ 2p - \left(p_{k-1}^+ + p_{k-1}^- \right) \right\}} \quad \leqslant \quad \frac{\sum \left\{ 2p - \left(p_{k-1}^+ + p_{k-1}^- \right) \right\}^{-1}}{\sum_{k=1}^{n} \tilde{\omega}_k(p)}$$

hence

$$\frac{\hat{J}_n^{3/2}(p)}{|\hat{J}_n'(p)|} \quad \leqslant \quad \frac{\sum \left\{ 2p - \left(p_{k-1}^+ + p_{k-1}^- \right) \right\}^{-1}}{\sqrt{\sum_{k=1}^{n} \tilde{\omega}_k(p)}}$$

Another Hölder's inequality gives $\dfrac{\hat{J}_n^{3/2}(p)}{|\hat{J}_n'(p)|} \leqslant \sqrt{\displaystyle\sum_{k=1}^{n} \frac{\left(p_{k-1}^+ - p\right)\left(p - p_{k-1}^-\right)}{2p - \left(p_{k-1}^+ + p_{k-1}^-\right)}}$ and simple calculus shows that each term of the sum is stricly smaller than 1. Then

$$|Z_n| \quad \leqslant \quad \sqrt{n} \left| \frac{\hat{J}_n(p)}{J_n(p)} - 1 \right| \quad \leqslant \quad \frac{2p^2}{\sqrt{n}} \sum_{k=1}^{n} \left(\tilde{\omega}_k(p) - \mathrm{E}\left[\tilde{\omega}_k(p)\right] \right)$$

from (3.7), then $Z_n \xrightarrow{I\!P} 0$ if (i) remains true $\forall \delta \geqslant 1/2$. □

8. Appendix B: supplementary Material

This supplementary section first provides details about the implementation of sweepline algorithms to solve Klee's measure problem, which allows for an exact computation (up to numerical rounding errors) of the probability bounds (p_n^-, p_n^+) ; a pseudo-code is given for direct use. Then a general result is given about the preservation of monotonicity when the uniform input $\mathbf{x} = (x_1, \ldots, x_d)$ results from an inverse transformation of the multivariate distributional transform.

8.1. A sweepline algorithm to compute volumes of hypercubic unions

Sweepline (or *plane sweep*) algorithms are commonly used to jointly detect and sort intersections between segments [39]. The d-dimensional volume is calculated recursively by exploring all n-1-dimensional "slices" of the d-th dimension. See [37, 12] and [7] for more explanations. When segments are parallel or perpendicular such as their intersections define a union of hypercubes sharing the same orthogonal basis, the volume calculation is known as Klee's measure problem [16, 5]. A pseudo-code follows to be used for direct implementation.

Let Δ_n be the $n \times d$ matrix of n vertices $(\mathbf{x}_1, \ldots, \mathbf{x}_n)$ defining the union of hypercubes (for an example, see Figure 1). In the following pseudo-code, the volume considered is V_n^-, also defined by the points of Δ_n and the origin $(0, \ldots, 0)$ of the $\mathbb{U}-$space.

Algorithm $\texttt{VOL}(\Delta_n, n, d)$.

1. Let $\Delta'_n = \sigma_{n,d}(\Delta_n)$ be the $n \times d$ permutation of Δ_n arranged in the increasing order of the $n-$vector of $d-$dimensional components.
2. Remove the $d-$dimensional components from Δ'_n and denote $\texttt{Vol}_n = 0$.
3. For $i \in \{1, \ldots, n\}$,
   (a) Consider the slice $\Delta_n^{(i)} = \left\{\mathbf{x}'_i, \ldots, \mathbf{x}'_n \in \Delta'_n\right\}$.
   (b) Denote $\widetilde{\texttt{Vol}}_n^{(i)}$ the $d-1-$dimensional volume of $\Delta_n^{(i)}$.
   If $\dim Z_n^{(i)} = 1$,
   - $\Delta_n^{(i)}$ is a $n-i+1-$vector and $\widetilde{\texttt{Vol}}_n^{(i)} = \max\{\mathbf{x} \in \Delta_n^{(i)}\}$;
   - force i to the index of this maximal component in Δ'_n;

   else $\widetilde{\texttt{Vol}}_n^{(i)} = \texttt{VOL}(\Delta_n^{(i)}, n-i+1, d-1)$.
   (c) Let $\Lambda_i = \Delta'_n[i,d] - \Delta'_n[i-1,d]$ the size of $\Delta_n^{(i)}$ (assuming $\Delta'_n[0,d] = 0$).
   (d) Compute $\texttt{Vol}_n^{(i)} = \Lambda_i \cdot \widetilde{\texttt{Vol}}_n^{(i)}$ the $d-$dimensional volume.
   (e) Update the total volume $\texttt{Vol}_n = \texttt{Vol}_n + \texttt{Vol}_n^{(i)}$.

In practice, this algorithm remains little used for dimension d larger than 2 or 3. This is not surprising because its complexity $C(n,d)$ (the number of runs for a d-dimensional hypervolume between n points) is $O(n^d)$. This appears when considering the first developments of $C(n,d)$:

$$\begin{aligned} C(n,d) &= \sum_{k=0}^{n-1} C(n-k,d-1) = \sum_{k=0}^{n-1}(k+1)C(n-k,d-2), \\ &= \sum_{k=0}^{n-1}\left(\sum_{p=0}^{k-1} p\right) C(n-k,d-3), \\ &= \sum_{k=0}^{n-1} \frac{(k+1)(k+2)}{2}\, C(n-k,d-3), \\ &\dots \end{aligned}$$

Note however than the fastest version of this algorithm, proposed in [33], runs in time $O(n^{d/2}\log n)$ for $d \geqslant 3$. An alternative approach was presented in [7] with the same asymptotic performance, although its exposition was restricted to dimensions 3 and 4. At the present time the computational difficulties raised by diminishing the cost still remain open problems, although some slight improvements have recently been brought in [5]. Some ideas of possible improvements could possibly come from a parallel with multi-objective optimization contexts (cf. Remark 1 in the article). Indeed, algorithms running in polynomial time $O(n^{k_1}d^{k_2})$ to compute hypervolume metrics of Pareto frontiers have already been proposed in [19].

8.2. Preservation of monotonicity through space transformation

Consider $\tilde{g}$ a monotonic function with physical input random vector $\mathbf{y} = (y_1, \dots, y_d)$ and denote T its multivariate distributional transform. The methodology proposed in the article applies using the transformed function $g = \tilde{g} \circ T^{-1}$, provided T^{-1} is an increasing function of independent uniform inputs $\mathbf{x} = (x_1, \dots, x_d)$. This is ensured when $(y_1, \dots, y_d)$ are independent, since $T^{-1} = (F_1^{-1}, \dots, F_d^{-1})$ where F_i is the ith marginal cdf. In dependent cases (and possibly when the physical inputs mix continuous and discrete distributions), the generalized Rosenblatt's transform [36] may be used if the inputs can be stochastically conditioned, namely they can be sorted to get the explicit writing of the joint cdf

$$F(y_1, \dots, y_d) = F_1(y_1) \prod_{i=2}^{d} F_{i|1,\dots,i-1}(y_i|y_1, \dots, y_{i-1}).$$

Under this assumption, next lemma provides an intuitive sufficient condition for T^{-1} to be an increasing function of all $x_i \sim \mathcal{U}[0,1]$.

LEMMA 8.1. — *Assume that for $i = 2, \dots, d$, there exists a mapping f_i and a set of (possibly random) parameters θ_i independent of $Y_1, \dots, Y_i$ such that:*

(i) $Y_i = f_i(Y_1, \dots, Y_{i-1}, \theta_i)$,

(ii) *f_i is a globally increasing function of $Y_1, \dots, Y_{i-1}$;*

then $T^{-1}(\mathbf{x})$ is an increasing function of $\mathbf{x}$.

Multivariate normal distributions are often selected as approximate ways to tackle the difficulties of assessing correlations between input physical parameters, and therefore deserve a particular interest in the field of computer experiments. If Chen [6] obtained general results about the preservation of monotonicity when these distributions are given under the form of Gaussian copulas, an immediate corollary of Lemma 8.1 is that any standard binormal input distribution with positive correlation coefficient μ ensures that $T^{-1}(\mathbf{x})$ is increasing. Indeed, $\mathbf{Y} = (Y_1, Y_2)$ where $Y_1 \sim \mathcal{N}(0,1)$ and $Y_2 = \mu Y_1 + \sqrt{1-\mu^2}\theta$ with $\theta \sim \mathcal{N}(0,1)$. A similar result can be found for the class of elliptical bivariate copulas.

Proof of Lemma 8.1. — Assume (i). $\forall t \in I\!R, \quad \forall k \in \{2, \dots, d\}$, denote $p^t_{\theta_i}(Y_1, \dots, Y_{i-1}) = P\left(f_i(Y_1, \dots, Y_{i-1}, \theta_i) < t | Y_1, \dots, Y_{i-1}\right)$. Then, $\forall z \in I\!R$, let $A^t_{Y_1,\dots,Y_{i-1}}(z)$ denote the event $\{p^t_{\theta_i}(Y_1, \dots, Y_{i-1}) \leqslant z\}$. By definition,

$$F^{-1}_{i|1,\dots,i-1}(z|Y_1, \dots, Y_{i-1}) \quad = \quad \inf\left\{t \in I\!R \,|P\left(A^t_{Y_1,\dots,Y_{i-1}}(z)\right) = 1\right\}.$$

Assuming (ii), $p^t_{\theta_i}(Y_1, \dots, Y_{i-1})$ is a decreasing function of $Y_1, \dots, Y_{i-1}$. Thus, given t, the occurence of event $A^t_{Y_1,\dots,Y_{i-1}}(y)$ similarly decreases. Necessarily t increases, hence the minimum value of all $t \in I\!R$ such that $\mathcal{P}(A^t_{Y_1,\dots,Y_{i-1}}(z)) = 1$ increases. Hence $F^{-1}_{i|1,\dots,i-1}$ is an increasing function of $Y_1, \dots, Y_{i-1}, \forall i \in \{2, \dots, d\}$. Since $Y_1 = F^{-1}(X_1)$ is obviously an increasing function of X_1, a simple recursive reasoning shows that $F^{-1}_{i|1,\dots,i-1}$ is an increasing function of $X_1, \dots, X_{i-1}$. The statement of the lemma follows. □

9. Acknowledgements

E. de Rocquigny (École Centrale de Paris) must be thanked to have drawn my attention to this topic. I am especially grateful to all members of the MRI/T56 group and A. Dutfoy (EDF Research & Development) and Prof. Fabrice Gamboa and Dr. Thierry Klein (Institut de Mathématiques de Toulouse), for their support during this work. Finally, I am thankful to an anonymous reviewer who deeply read this article and made comments, critics and useful suggestions which greatly helped to improve it.

Bibliography

[1] Anonymous. — Open TURNS version 0.14.0 – Reference guide. Tech. rep., EDF-EADS-PhiMeca (2011).

[2] Bercu (B.). — Inégalités exponentielles pour les martingales [in french]. Journées ALEA 2008 CIRM p. 10-14 March (2008).

[3] Bester (C.), Hansen (C.). — Bias reduction for Bayesian and frequentist estimators. Working Paper. University of Chicago (2005).

[4] Cannamela (C.), Garnier (J.), Iooss (B.). — Controlled stratification for quantile estimation. Annals of Applied Statistics 2, p. 1554-1580 (2008).

[5] Chan (T.). — A (slightly) faster algorithm for Klee's measure problem. p. 94-100, College Park, MD, USA (2008).

[6] Chen (G.). — Monotonicity of dependence concepts: from independent random vector into dependent random vector. World Academy of Science, Engineering and Technology 57, p. 399-408 (2009).

[7] Chlebus (B.). — On the Klee's measure problem in small dimensions. Proceedings of the 25th Conference on Current Trends in Theory and Practice of Informatics (1998).

[8] Cox (D.), Snell (E.). — A general definition of residuals. Journal of the Royal Statistical Society 30, p. 248-275 (1968).

[9] Crowder (M.). — Maximum likelihood estimation for dependent observations. Journal of the Royal Statistical Society 38, p. 43-53 (1975).

[10] Crowder (M.). — On constrained maximum likelihood estimation with non-iid. observations. Annals of the Institute of Statistical Mathematics 36, p. 239-249 (1983).

[11] Daniels (H.), Velikova (M.). — Monotone and partially monotone neural networks. IEEE Transactions in Neural Networks 21, p. 906-917 (2010).

[12] de Berg (M.), van Kreveld (M.), Overmars (M.), Schwarzkopf (O.). — Computational Geometry Algorithms and Applications. Springer-Verlag (1997).

[13] de Rocquigny (E.). — 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, p. 363-374 (2009).

[14] Durot (C.). — Monotone nonparametric regression with random design. Mathematical Methods in Statistics 17, p. 327-341 (2008).

[15] Efron (B.), Tibshirani (R.J.). — An Introduction to the Bootstrap. New York: Chapman & Hall (1993).

[16] Erickson (J.). — Klee's measure problem. Tech. rep., University of Illinois at Urbana-Champaign, URL: http://theory.cs.uiuc.edu/jeffe/open/klee.html (1998).

[17] Ferrari (S.), Cribari-Neto (F.). — On bootstrap and analytical bias correction. Economics Letters 58, p. 7-15 (1998).

[18] Figueira (J.), Greco (S.), Erhgott (M.). — Multiple criteria decision analysis – State of the art – Survey. Springer's International Series (2005).

[19] Fleischer (M.). — The measure of Pareto optima. Applications to multi-objective metaheuristics. vol. 262, p. 519-523, Faro, Portugal (2003).

[20] Glynn (P.), Rubino (G.), Tuffin (B.). — Robustness properties and confidence interval reliability. In: Rare Event Simulation. Wiley (2009).

[21] Hurtado (J.). — An examination of methods for approximating implicit limit state functions from the viewpoint of statistical learning theory. Structural Safety 26, p. 271-293 (2004).

[22] Kleijnen (J.). — Simulation optimization via bootstrapped kriging: Survey. Tech. rep., Rport from Tilburg University, Center for Economic Research (2011).

[23] Kleijnen (J.), van Beers (W.). — Monotonicity-preserving bootstrapped kriging metamodels for expensive simulations. Tech. rep., Discussion Paper 2009-75, Tilburg University, Center for Economic Research (2009).

[24] Kroese (D.), Rubinstein (R.). — Simulation and the Monte Carlo Method (2nd edition). Wiley (2007).

[25] Lemaire (M.), Pendola (M.). — PHIMECA-SOFT. Structural Safety 28, p. 130-149 (2006).

[26] Limbourg (P.), de Rocquigny (E.), Andrianov (G.). — Accelerated uncertainty propagation in two-level probabilistic studies under monotony. Reliability Engineering and System Safety 95, p. 998-1010 (2010).

[27] Lin (D.). — A new class of supersaturated design. Technometrics 35, p. 28-31 (1993).

[28] MacKay (M.), Beckman (R.), Conover (W.). — A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. Technometrics 21, p. 239-249 (1979).

[29] Madsen (H.), Ditlevsen (O.). — Structural reliability methods. Wiley (1996).

[30] Meyer (P.A.). — Martingales and stochastic integrals. Lecture Notes in Mathematics Vol. 284. Springer-Verlag (1972).

[31] Morio (J.). — Influence of input pdf parameters of a model on a failure probability estimation. Simulation Modelling Practice and Theory 19, p. 2244-2255 (2011).

[32] Munoz-Muniga (M.), Garnier (J.), Remy (E.), de Rocquigny (E.). — Adaptive directional stratification for controlled estimation of the probability of a rare event. Reliability Engineering and System Safety (in press) (2011).

[33] Overmars (M.H.), Yap (C.K.). — New upper bounds in Klee's measure problem. SIAM Journal of Computing 20, p. 1034-1045 (1991).

[34] Rajabalinejad (M.), Meester (L.), van Gelder (P.), Vrijling (J.). — Dynamic bounds coupled with Monte Carlo simulations. Reliability Engineering and System Safety 96, p. 278-285 (2011).

[35] Ranjan (P.), Bingham (D.), Michailidis (G.). — Sequential experiment design for contour estimation from complex computer codes. Technometrics 50, p. 527-541 (2008).

[36] Rüschendorf (L.). — On the distributional transform, Sklar's theorem, and the empirical copula process. Journal of Statistical Planning and Inference 139, p. 3921-3927. (2009)

[37] Shamos (M.I.), Hoey (D.). — Geometric intersection problems. Proceedings of the 17th IEEE Symposium about the Foundations of Computer Science (FOCS'76) p. 208-215 (1976).

[38] Tsompanakis (Y.), Lagaros (N.), Papadrakakis (M.), Frangopol (D.e.). Structural design optimization considering uncertainties. Taylor & Francis (2007).

[39] van Leeuwen (J.), Wood (D.). — The measure problem for rectangular ranges in d-space. Journal of Algorithms 2, p. 282-300 (1981).