Bounding and estimating an exceedance probability in output from monotonous computer codes

Nicolas Bousquet

arXiv:1012.1042v6 [math.ST] 21 Dec 2010

Abstract This article deals with the estimation of a probability p_f of an undesirable event. Its occurence is formalized by the exceedance of a threshold reliability value by the unidimensional output of a computer code G with multivariate probabilistic input $\mathbf{X}$. When G is assumed time-consuming and monotonous with respect to $\mathbf{X}$, the Monotonous Reliability Method, proposed by de Rocquigny (2009) in an engineering context, can provide bounds and crude estimates of p_f, via deterministic or stochastic designs of experiments. The present article consists in a formalization and technical deepening of this idea, as a large basis for future theoretical and applied studies. Three kinds of results are especially emphasized. First, the bounds themselves remain too crude and conservative estimators of p_f for a dimension of X upper than 2. Second, a maximum-likelihood estimator of p_f can be easily built, presenting a high variance reduction with respect to a standard Monte Carlo case, but suffering from conservative bias. Third, the theoretical properties of a family of unbiased estimators of p_f, based on sequential nested importance samplings, are analyzed. Their supplementary potential improvement requires further studies whose main lines are discussed. Along the paper, the efficiency and difficulties of these approaches are illustrated by a generic example. In fine, we show that both approaches lead to promising parsimonious estimation algorithms provided a sequential emulation of the limit state (failure) surface, seen as a supervised classification problem, can be made under monotony constraints. Besides, some connections and research avenues are identified in various mathematical areas like multivariate statistics, multi-objective optimization and computational geometry.

N. Bousquet
EDF R&D, 6 quai Watier, 78401 Chatou, France.
Tel.: +33-01-30-87-75-08
Fax: +33-30-87-82-13
E-mail: nicolas.bousquet@edf.fr

Keywords Exceedance probability · Computer code · Input uncertainty · Monte Carlo acceleration

1 Introduction

In many technical areas, the exceedance of some unidimensional variable Z over a certain critical value z^* is an event the probability p_f of which must be carefully monitored. For instance, a conservative estimation of a river flood probability p_f, defined by the downstream water height Z and a given dyke height z^*, can be a fundamental task required by control authorities. The term *conservative* means here that p_f should not be underestimated. Since the natural framework of such a concern is structural reliability, where the event is often undesirable, p_f will be often called the *failure* probability and obviously be assumed stricly positive.

We consider here the frequent case where $z^* - Z = G(\mathbf{X})$ where G is a deterministic function, usually a computer code, and $\mathbf{X}$ is a vector of uncertain input parameters. In the flooding example, $\mathbf{X}$ can typically include parameters of river geometry, friction coefficients and upstream rainfall height, and G is a hydraulical code resolving numerically fluid mechanism equations. Probabilistic approaches being commonly used to take account of the variability of input parameters, $\mathbf{X}$ is assumed to be a random vector with probability density function (pdf) $f_{\mathbf{X}}$ in a $d-$dimensional space $\mathbb{D}$.

The most traditional way to estimate

$$p_f = P(G(\mathbf{X}) \leq 0) \; = \; \int_{\mathbb{D}} \mathbb{1}_{\{G(\mathbf{x})\leq 0\}} f_{\mathbf{X}}(\mathbf{x}) \, d\mathbf{x}.$$

is adopting a Monte Carlo (MC) strategy, namely p_f is estimated by $\hat{p_f}_n = n^{-1} \sum_{k=1}^{n} \mathbb{1}_{\{G(\mathbf{X}_k)\leq 0\}}$ where n is large and the $\mathbf{X}_k$ are independently sampled according to $f_{\mathbf{X}}$. This estimator is well known to present good convergence properties and does not require any regularity hypotheses on G or the limit state surface (or *failure surface*) $\mathbb{D}_l = \{\mathbf{x} \in \mathbb{D} \,|G(\mathbf{x}) = 0\}$. Furthermore the estimation precision is independent on

the dimension d. It is thus the most appropriate method when G is a complete black box.

This strategy presents however some difficulties which prohibit its use in many problems, especially those for which a trial of G is very time-consuming. Indeed, a good estimation of a low probability $p_f \sim 10^{-q}$ typically requires at least 10^{q+2} trials (Lemaire and Pendola 2006). Furthermore $\hat{p_f}_n$ is not a robust estimator. In the sense given by Glynn et al. (2009), it means that for any n its relative error, namely its coefficient of variation $\mathrm{CV}[\hat{p_f}_n]$, tends to an infinite limit when $p_f \to 0$.

Many non-intrusive strategies have been proposed to accelerate the MC approach and build parsimonious estimators, the variances of which being smaller than the variance $V^{MC} = p_f(1-p_f)/n$ of the MC estimator. Traditional methods of the engineer community in structural reliability (First and Second-Order Reliability Methods (FORM/SORM), cf. Madsen and Ditlevsen 1996) consider the estimation of p_f as an optimization rather than a propagation problem. Those methods are generally very parsimonious but lead to estimators with weakly or non-controllable error.

Statistical methods like quasi-MC, sequential MC or importance sampling approaches (Kroese and Rubinstein 2007) are based on a static or sequential selection of the *design of experiments* (DoE), namely the set of points $\mathbf{x_k}$ on which G is tried, to improve the covering of $\mathbb{D}$ in areas close to the limit state surface. Most advanced methods often get rid of the time-consuming difficulties by emulating the behavior of G (Cannamela et al. 2008), for instance using kriging techniques. Although they introduce prediction error, such techniques may appear necessary to implement Markovian particle-based strategies (L'Ecuyer et al. 2007) or *subset* simulations (Au and Beck 2001), which build sequences of conditional estimations of p_f and can lead to robust estimators of very low probabilities.

All these methods try to minimize the strenght of the hypotheses placed on G. However, in the engineering practice the behavior of Z can be known to be monotonous with respect to $\mathbf{X}$. In the flooding example, an increasing of the upstream rainfall usually implies an increasing of the downstream water height toward a certain limit linked to the topological features of the river. See de Rocquigny (2009) for further details on this example and Rajabalinejad et al. (2010) on a similar one. Monotonous properties of computer codes have been considered in various theoretical and engineering domains, e.g. proving the MC acceleration of Latin Hypercube Sampling (MacKay et al. 1979) for the estimation of expectancies, carrying out screening methods for sensitivity analyses (Lin 1993), predicting the behavior of network queuing systems (Ranjan et al. 2008) or estimating the safety of a nuclear reactor pressure vessel (Munoz-Muniga et al. 2010). Given $\mathbf{X}$ and assuming $z^* - Z(\mathbf{X})$ can reliably be modelled by $G(\mathbf{X})$, the code monotony itself is assumed. Most traditional relaxations of this hypothesis are assuming $Z = g(\mathbf{X}, \varepsilon)$ where $g(\mathbf{x}, \varepsilon) = G(\mathbf{x}) + \varepsilon$ or $g(\mathbf{x}, \varepsilon) = G(\mathbf{x})\varepsilon$ where ε is some *model error*. Provided ε can be assessed, g is still a monotonous code.

Recently some articles (de Rocquigny 2009, Limbourg et al. 2010, Rajabalinejad et al. 2010) highlighted the possibility of taking advantage from code monotony to bound and estimate p_f. Under the name of *Monotonous Reliability Methods* (MRM), De Rocquigny (2009) proposed a class of sequential algorithms which build a progressive bounding of the limit state surface, thanks to a sequential DoE, which allows for the computation of a crude estimator of p_f at each step. Mainly aimed at the community of engineers in structural reliability, these articles consisted in a brief description of the MRM features and behavioral studies on some examples. On the case-studies they considered, their empirical results showed that the precision of FORM/SORM methods, as well as a classic MC approach (for a same number of trials), was significantly improved. Similar results were obtained by Rajabalinejad et al. (2010). However, although a parallelization of such algorithms was already implemented (Limbourg et al. 2010), the convergence properties of the MRM class have not been theoretically studied yet.

The aim of the present article is to start addressing this issue. Our main contribution is a clear formalization of MRM, the proper definition of two classes of estimators of p_f, and providing some theoretical conditions on the DoE and the limit state surface such that these estimators have better properties than MC ones. A second contribution is to offer a practical view of the technical tools needed to study the sequential inference techniques based on code monotony. Finally, along the paper some connections are done with other areas of computational mathematics, especially about implementation issues.

To go more into details, we describe in Section 2 the general principle of MRM algorithms. Section 3 deals with a first estimator of p_f, defined as the maximum likelihood of progressively sampled dependent-data, for which asymptotic normality conditions are provided, implying robustness, conservatism and a theoretically good MC acceleration. Furthermore it does not require any calibration work. This estimator is however built on a sequence of nested naive uniform sampling, and can suffer from high bias when the dimension increases. Therefore Section 4 is dedicated to the study of a class of unbiased estimators based on nested generalized importance sampling. Their study shows that a significant improvement of the first results emanates from the sequential resolution of a supervised classification problem under monotony constraints. A Discussion section ends this

paper, focusing especially on the research avenues that must be explored to improve the theoretical and applied results presented here, before getting turnkey estimators. Guidelines are also provided to connect the general convergence results presented here with advanced techniques evoked hereinbefore. Finally, note that the technical proofs of all statements in the main text are given in the Appendix.

2 The principle of Monotonous Reliability Methods

2.1 Working assumptions, definitions and basic properties

Let $\mathbb{D} : \mathbf{X} \mapsto G(\mathbf{X})$ be a deterministic computer code defined as a real-valued scalar function of $\mathbf{X} = (X_1, \ldots, X_d)$ on its definition domain $\mathbb{D} \subset \mathbb{R}^d$. By *deterministic*, we mean that $G(\mathbf{x})$ produces the same output every time if it is given the same input $\mathbf{x}$. A probability space $(\mathbb{D}, \mathscr{A}, P)$ is defined in order to model the input uncertainty, giving to $\mathbf{X}$ the nature of a random vector with joint pdf $f_{\mathbf{X}}$ and cumulative distribution function (cdf) $F_{\mathbf{X}}$.

Assumption 1. G is *globally increasing* over $\mathbb{D}$.

Global monotony can be defined as follows: $\forall i$, $\exists s_i \in \{-1,+1\}$, $\forall \varepsilon > 0$, $\forall \mathbf{x} = (x_1, \ldots, x_d) \in \mathbb{D}$, such that

$$G(x_1, \ldots, x_{i-1}, x_i + s_i\varepsilon, x_{i+1}, \ldots, x_d) \quad \leq \quad G(x_1, \ldots, x_{i-1}, x_i, x_{i+1}, \ldots, x_d)$$

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. In this paper, the increasing monotony assumption is made without loss of generality since any decreasing $i-$th component can be changed from x_i to $-x_i$.

Assumption 2. G is continuous with respect to all its inputs (possibly extended to continuous domains).

The smoothness of G (ie. $\mathbf{x} \mapsto G(\mathbf{x})$ is differentiable) is not required for the algorithms proposed by De Rocquigny (2009) and Limbourg et al. (2010). This relaxes a standard hypothesis placed on computer code functions when dealing with estimations based on interpolation (O'Hagan 2006). However, further in the text a smoothness assumption on the limit state surface is needed to get theoretical convergence results on the estimation of p_f.

Definition 1 A set of points of $\mathbb{D}$ is said *safety-dominated* (*resp. failure-dominated*) if G is guaranteed to be positive (*resp.* negative) in any point of this set.

Denote by the symbol $\succeq$ the partial order between elements of any $d-$dimensional space in bijection with $\mathbb{D}$, namely $\mathbf{x} \succeq \mathbf{y}$ means that all components of $\mathbf{x}$ are together larger or equal to the components of $\mathbf{y}$. Then assume that some point value $G(\tilde{\mathbf{x}})$ is known, and consider the sets

$$\mathbb{D}^+_{\tilde{\mathbf{x}}} = \{\mathbf{x} \in \mathbb{D} \mid \mathbf{x} \succeq \tilde{\mathbf{x}}\}, \quad \mathbb{D}^-_{\tilde{\mathbf{x}}} = \{\mathbf{x} \in \mathbb{D} \mid \mathbf{x} \preceq \tilde{\mathbf{x}}\}.$$

The increasing monotony implies that if $G(\tilde{\mathbf{x}}) > 0$ (*resp.* $G(\tilde{\mathbf{x}}) < 0$), then $\mathbb{D}^+_{\tilde{\mathbf{x}}}$ is safety-dominated (*resp.* $\mathbb{D}^-_{\tilde{\mathbf{x}}}$ is failure-dominated). This proves the next lemma.

Lemma 1 *Both inequalities are true with probability 1:*

$$p_f \leq 1 - P(\mathbf{X} \in \mathbb{D}^+_{\tilde{\mathbf{x}}}) \quad \textit{if } G(\tilde{\mathbf{x}}) > 0,$$
$$p_f \geq P(\mathbf{X} \in \mathbb{D}^-_{\tilde{\mathbf{x}}}) \quad \textit{else.}$$

More generally, if n trials of the numerical model G have been performed for a sample of n input vectors $(\mathbf{x}_j)_{j=1,\ldots,n}$, we may group them into the safe and failure sub-samples following the corresponding values of $(G(\mathbf{x}_j))_{j=1,\ldots,n}$, respectively

$$\Xi^+_n = \left\{\mathbf{x} \in (\mathbf{x}_j)_{j=1,\ldots,n} \mid G(\mathbf{x}_j) \geq 0\right\},$$
$$\Xi^-_n = \left\{\mathbf{x} \in (\mathbf{x}_j)_{j=1,\ldots,n} \mid G(\mathbf{x}_j) \leq 0\right\},$$

then generate the sets

$$\mathbb{D}^+_n = \left\{\mathbf{x} \in \mathbb{D} \mid \exists \mathbf{x}_j \in \Xi^+_n,\ \mathbf{x} \succeq \mathbf{x}_j\right\},$$
$$\mathbb{D}^-_n = \left\{\mathbf{x} \in \mathbb{D} \mid \exists \mathbf{x}_j \in \Xi^-_n,\ \mathbf{x} \preceq \mathbf{x}_j\right\}.$$

Doing so we obtain generalized bounds for p_f: denoting $p^-_n = P(\mathbf{X} \in \mathbb{D}^-_n)$ and $p^+_n = 1 - P(\mathbf{X} \in \mathbb{D}^+_n)$ to alleviate the notations, we have

$$p^-_n \leq p_f \leq p^+_n. \tag{1}$$

Hereafter, $\mathbb{D}^+_n$ and $\mathbb{D}^-_n$ will be referred to as *dominated spaces*, where the sign of $G(\mathbf{X})$ is known. Note that the complementary *non-dominated* space

$$\mathbb{D}_n = \mathbb{D} / \left(\mathbb{D}^+_n \oplus \mathbb{D}^-_n\right)$$

is the only part of space where further trials of G are required to improve the bounding and some estimation of p_f.

Precision criterion. Assuming a computational budget independent on p_f, it is desirable to evaluate the precision reached by the computation along the algorithmic steps (crudely speaking, the width between the bounds). Various criteria $\rho_n = \rho(p^-_n, p^+_n)$ can be considered according to the context of the study. For instance, typical safety studies in the nuclear field would categorize failure or initiating events according to logarithmic probability classes. In a concern of generality, a criterion built on relative bounding precision appears desirable, e.g.

$$\Upsilon_n(p_f) = (p^+_n - p^-_n)/p_f = |p^+_n/p_f - 1| + |1 - p^-_n/p_f|.$$

Reaching a precision order ε on $\Upsilon_n(p_f)$, independently of the unknown p_f, also appears simple. It is enough to run MRM until

$$\rho_n = \Upsilon_n(p_n^-) \;=\; (p_n^+ - p_n^-)/p_n^- \;\leq\; \varepsilon$$

since one always has $\Upsilon_n(p_f) \leq \rho_n$.

Remark 1 In multi-objective optimization (MOO ; Figueira et al. 2005), a dominated space can be interpreted as a subset of a performance space delimited by a Pareto frontier. In this framework, the code is thought as a monotonous rule of decision depending of d variables, for which the set of n best possible configurations (the frontier) is searched. Both frameworks share concerns of parsimony because of the limited amount of time in practice for examining the applicability of the solutions in MOO.

2.2 The algorithm

2.2.1 Space transformation

Transforming the input space often appears as a preliminary task of well-known methods in structural reliability to simplify its exploration. For instance FORM/SORM methods involve the two-steps Nataf iso-probabilistic transformation of $\mathbb{D}$ to a standardized Gaussian space. The rationale for such transformations is that estimating p_f appears simpler. In our case, a link can be done between the computation of the bounds around p_f and a classic computational geometry problem which can be exactly solved, as explained in Appendix B.

Assumption 3. There exists an invertible transformation $\Psi_{F_\mathbf{X}} : \mathbb{D} \mapsto \mathbb{U}$ where $\mathbb{U}$ is the $d-$dimensional hypercube $[0,1]^d$, such that:

(1) the transformed input vector $\boldsymbol{\nu} = \Psi_{F_\mathbf{X}}(\mathbf{X})$ has all its components independent and identically distributed (iid) and follows a uniform distribution on $\mathbb{U}$, the new failure function (or code) G becoming

$$\tilde{G}(\boldsymbol{\nu}) = G \circ \Psi_{F_\mathbf{X}}^{-1}(\boldsymbol{\nu}) \;=\; G(\mathbf{X});$$

(2) $\tilde{G}$ remains a monotonic (increasing) function of the new input vector $\boldsymbol{\nu}$.

When the X_i are independent, this assumption is always satisfied since $\Psi_{F_\mathbf{X}}^{-1}$ can simply be chosen as the product of the marginal cdfs. In dependent cases, explanations and some technical requirements about the choice of $\Psi_{F_\mathbf{X}}^{-1}$ are given in Appendix A. The image space is now described as the probability space $(\mathbb{U}, \mathscr{B}(\mathbb{U}), P)$.

2.2.2 The algorithmic scheme

Denote now $\mathbb{U}_n^-$, $\mathbb{U}_n^+$ and $\mathbb{U}_n = \mathbb{U}/(\mathbb{U}_n^+ \oplus \mathbb{U}_n^-)$ the image spaces through $\Psi_{F_\mathbf{X}}^{-1}$ of $\mathbb{D}_n^-$, $\mathbb{D}_n^+$ and $\mathbb{D}_n$. The scheme shared by all MRM variants can be described as follows.

MRM Global Scheme

Step 0. `Initialization:`
- (a$_0$) `let` $\mathbb{U}_0^+ = \{1^d\}$, $\mathbb{U}_0^- = \{0^d\}$ `and` $\mathbb{U}_0 = [0,1]^d$
- (b$_0$) `let` $(p_0^-, p_0^+) = (0,1)$
- (c$_0$) `select an initial DoE` $\{\boldsymbol{\nu}_1^{(1)}, \ldots, \boldsymbol{\nu}_1^{(m_1)}\}$

Step $n \geq 1$. `While` $\rho(p_{n-1}^-, p_{n-1}^+) > \varepsilon$
- (a) **if** $n > 1$, `select a DoE` $\{\boldsymbol{\nu}_n^{(1)}, \ldots, \boldsymbol{\nu}_n^{(m_n)}\} \in \mathbb{U}_{n-1}$
- (b) `compute the` *signatures*

$$\xi_{\boldsymbol{\nu}_n}^{(j)} = \mathbb{1}_{\left\{\tilde{G}\left(\boldsymbol{\nu}_n^{(j)}\right) < 0\right\}} \tag{2}$$

- (c) `update the spaces`

$$\mathbb{U}_n^- = \mathbb{U}_{n-1}^- \cup \mathbb{U}^-$$
$$\mathbb{U}_n^+ = \mathbb{U}_{n-1}^+ \cup \mathbb{U}^+$$
$$\mathbb{U}_n = \mathbb{U}/(\mathbb{U}_n^- \cup \mathbb{U}_n^+)$$

`where`

$$\mathbb{U}^- = \left\{\boldsymbol{\nu} \in \mathbb{U} \;\middle|\; \exists\; \boldsymbol{\nu}_n^{(j)},\quad \xi_n^{(j)} = 1,\quad \boldsymbol{\nu} \preceq \nu_n^{(j)}\right\}$$
$$\mathbb{U}^+ = \left\{\boldsymbol{\nu} \in \mathbb{U} \;\middle|\; \exists\; \boldsymbol{\nu}_n^{(j)},\quad \xi_n^{(j)} = 0,\quad \boldsymbol{\nu} \succeq \boldsymbol{\nu}_n^{(j)}\right\}$$

`and their volumes` (V_n^-, V_n^+)
- (e) `update the bounds` $(p_n^-, p_n^+) = (V_n^-, 1 - V_n^+)$

At each step, the DoE must be chosen taking account of the increasing monotony of $\tilde{G}$. Denoting $\boldsymbol{\nu}_n^{(1)}$ and $\boldsymbol{\nu}_n^{(2)}$ two points of the DoE and assuming to know $\xi_n^{(1)}$, the signature of $\boldsymbol{\nu}_n^{(2)}$ is unnecessary to compute in two cases:

if $\xi_n^{(1)} = 1$ and $\boldsymbol{\nu}_n^{(1)} \succeq \boldsymbol{\nu}_n^{(2)} \Rightarrow \boldsymbol{\nu}_n^{(2)} \in \mathbb{U}^-_{\boldsymbol{\nu}_n^{(1)}}$ and $\xi_n^{(2)} = 1$,

if $\xi_n^{(1)} = 0$ and $\boldsymbol{\nu}_n^{(1)} \preceq \boldsymbol{\nu}_n^{(2)} \Rightarrow \boldsymbol{\nu}_n^{(2)} \in \mathbb{U}^+_{\boldsymbol{\nu}_n^{(1)}}$ and $\xi_n^{(2)} = 0$.

Thus the order of $\tilde{G}$ trials should be carefully monitored, in relation with the partial order between the points in the DoE. Reducing the DoE to a single point, i.e. $m_n = 1$ for all steps, minimizes the number of unnecessary trials. This strategy is favored in the present paper. A two-dimensional example of MRM progression is displayed on Figure 1.

2.2.3 Initialization via deterministic Design of Experiments

First MRM iterations should be monitored to reduce significantly the width of interval $[p_n^-, p_n^+]$, such that later iterations mainly focus on scale refinements. When p_f corresponds to the probability of a rare event, one could hope that

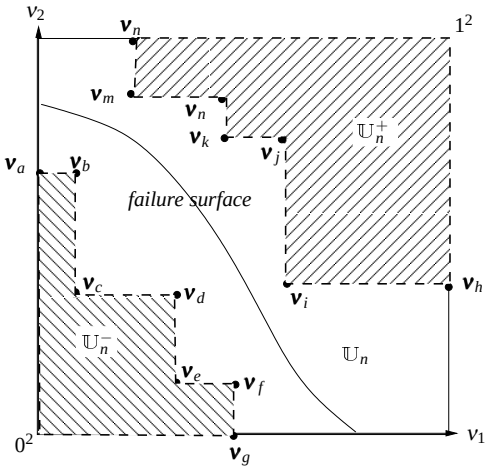

Fig. 1 Two-dimensional uniform space after $n = 14$ MRM iterations. Points $\{0^2, \boldsymbol{v}_a, \boldsymbol{v}_b, \boldsymbol{v}_c, \boldsymbol{v}_d, \boldsymbol{v}_e, \boldsymbol{v}_f, \boldsymbol{v}_g\}$ have positive signatures and are vertexes of $\mathbb{U}_n^-$. Points $\{\boldsymbol{v}_h, \boldsymbol{v}_i, \boldsymbol{v}_j, \boldsymbol{v}_k, \boldsymbol{v}_l, \boldsymbol{v}_m, \boldsymbol{v}_n, 1^2\}$ have zero signatures and are vertexes of $\mathbb{U}_n^+$.

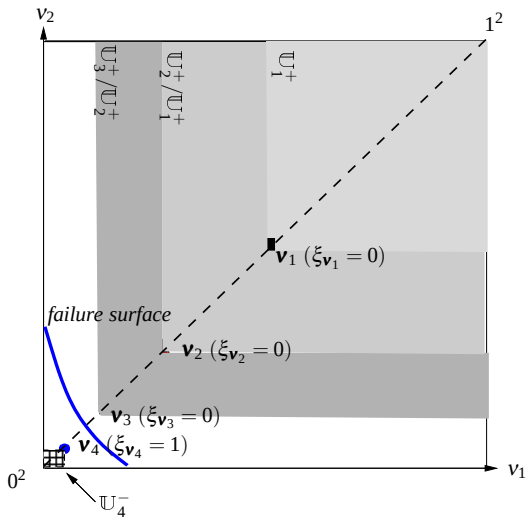

Fig. 2 Diagonal deterministic (DD-MRM) strategy, assuming a low p_f, stopping after 4 steps.

p_n^+ rapidly evolves from 1 to a low value. Most interesting refinements begin when $p_n^- > 0$, namely when one finds the first point with signature 1 other than the origin 0^d. In this perspective, we suggest to start the algorithm using a deterministic DoE until reaching $p_n^- > 0$, then switch on another DoE, typically *stochastic* (cf. § 2.2.4).

De Rocquigny (2009) suggested using isoprobabilistic DoEs gridding regularly $\mathbb{U}$. Such algorithms, however, involve exponentially-increasing numbers of trials, which upholds the need for switching to another DoE providing fastest explorations of the uniform space. In the applications presented further, a dichotomic diagonal MR method (DD-MRM), illustrated on Figure 2 in a 2-dimensional case, was used. It explores the non-dominated space $\mathbb{U}_n$ in a very intuitive way, maximizing the removable volume at each iteration, and stops at step $k_0 \geq 1$ such that

$$k_0 \geq \frac{\log(1/p_f)}{d \log 2}.$$

Consequently, an expected crude prior value of p_f can help to estimate the minimal number k_0 of trials.

2.2.4 Stochastic Designs of Experiments

From now and to the end of the main text, without loss of generality, we assume to start the notation $(\mathbb{U}_0^+, \mathbb{U}_0^-, p_0^+, p_0^-)$ after $N-1$ introductive deterministic steps with $N \geq k_0 + 1$. From step N, the DoE is chosen stochastic and we denote $\mathscr{F}_n$ the $\sigma-$algebra generated by a series of n stochastic samplings $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n$ (although this notation is sometimes dismissed to alleviate the text when there cannot be misunderstanding). All stochastic processes considered in the following are adapted to the filtration $(\mathscr{F}_n)_n$.

A stochastic exploration of succesive non-dominated spaces has two advantages. First, it is likely to diminish the cost of deterministic DoEs. Second, it allows for a statistical estimation of p_f, which is what the remainder of the paper deals with. Before that, some basic observations can be done about the behavior of the bounds and the precision criterion.

Obviously, $1 - p_n^-$ and p_n^+ both decrease and are bounded in $[0,1]$. Since p_0^- and p_0^+ are known, these processes are *predictible*.

Thus $1 - p_n^-$ and p_n^+ are supermartingales bounded in $\mathbb{L}^p$ $\forall p \geq 1$. Therefore, from generalized Doob's theorem (Bercu 2008), there exists two random limits (p_∞^-, p_∞^+) such that $0 < p_\infty^- \leq p_f \leq p_\infty^+ < 1$ and $p_n^- \xrightarrow{a.s., \mathbb{L}^p} p_\infty^-$ and $p_n^+ \xrightarrow{a.s., \mathbb{L}^p} p_\infty^+$. The sequence of random variables $\{\rho_n\}$ is a strictly decreasing predictive process, and converges similarly toward the random variable $\rho_\infty = p_\infty^+ / p_\infty^- - 1 \geq 0$. Lebesgue's monotonous convergence theorem implies almost sure convergence of expectancies of p_n^-, p_n^+ and ρ_n conditioned on $\mathscr{F}_{n-k}$. Using Theorem 2.1 in Cadre (2002), the sequence $\{\rho_n^{\text{med}}\}$ of median values of ρ_n along a sampling strategy is decreasing and converges uniformly to $\text{Med}[\rho_\infty]$.

Example 1 Along the paper, the results will be mostly illustrated using the following generic toy example. For a given dimension d, let

$$Y_d = H_d(\mathbf{X}) \ = \ X_1 / (X_1 + \sum_{i=2}^{d} X_i)$$

where the input X_i follows the gamma distribution $\mathscr{G}(i+1, 1)$. Obviously, $\forall\, d \geq 2$, G_d is increasing in $(-X_1, X_2 \ldots, X_d)$ and Y_d follows the beta distribution $\mathscr{B}e(2, 2^{-1}(d+1)(d+2)-$

3). Therefore, denoting q_{d,p_f} the p_f-order quantile of Y_d, the code defined by

$$G_d(\mathbf{X}) = H_d(\mathbf{X}) - q_{d,p_f}$$

is related to the known exceedance probability p_f.

Choosing $p_f = 5\%$ and a sampling strategy of the DoE only based on nested uniform sampling, the behavior of the bounds is displayed on Figure 3 for dimensions 2 to 4, in addition to this of the 95% confidence interval of a standard MC estimation. When $d = 2$, the "100%-confidence" deterministic bounds lead to a significantly better precision sharper than the statistical confidence interval. The behavior of ρ_n and ρ_n^{med} is displayed on Figure 4.

Unidimensional cases. When $d = 1$, then $p_\infty^- = p_\infty^+ = p_f$ and the dynamic bounds are consistent estimators of p_f, whatever the sampling of the DoE. Thus any estimator of p_f located between these bounds is automatically consistent, and $\rho_n \xrightarrow{a.s.} 0$ and $\rho_n^{\text{med}} \xrightarrow{\text{unif.}} 0$.

Indeed, it is enough to notice that the sequence $a_n = |p_n^- - p_f| = p_f - p_n^-$ is decreasing. Therefore one can extract a strictly decreasing sequence $\{a_{\sigma(n)}\}$ from $\{a_n\}$. Since $a_{\sigma(n+1)}/a_{\sigma(n)} < 1$, the sum $\sum_{k=1}^{n} a_{\sigma(k)}$ is converging, which implies $a_{\sigma(n)} \to 0$ then $a_n \to 0$. Then $\exists c_n \to 0$ such that $|p_n^- - p_f| \leq c_n$, which proves $p_n^- \xrightarrow{a.s.} p_f$.

Multidimensional cases. In multidimensional cases, we conjecture[1] that the limits p_∞^- and p_∞^+ may be not equal to p_f. This counter-intuitive idea results from the following speculation: in dimensions higher than 2, the addition of hypercubes with diminishing sizes seems not to asymptotically fill the non-dominated space if the limit state surface presents some linearity in a non-empty measurable set of points in $\mathbb{U}$. This speculation is precised in the following example, which was kindly provided by G. Bouchard (Xerox Research Centre Europe).

Example 2 Consider the particular case where the failure surface $\mathbb{U}_l$ is the diagonal hyperplane such that $p_f = 1/2$ $\forall d \geq 2$. Then we conjecture that any deterministic or stochastic $n-$sized DoE cannot be faster to cover $\mathbb{U}^-$ than the (fractalized) dichotomic strategy illustrated in Figure 5: at each step $k \geq 1$ with $k \leq \log_2(n+2)$, d^{k-1} hypercubes of side $(1/2)^k$, namely of volume 2^{-kd}, are built such that their higher vertice (the furthest to the origin) belongs to $\mathbb{U}_l$. When $n \to \infty$, the total dominated volume is

$$V_\infty(d) = \lim_{n\to\infty} \sum_{k=1}^{\log_2(n+2)} d^{k-1} 2^{-kd} = \frac{1}{d}\left(\frac{1}{1-2^{-d}d} - 1\right).$$

[1] This conjecture is maybe false and the rationale used here needs to be checked.

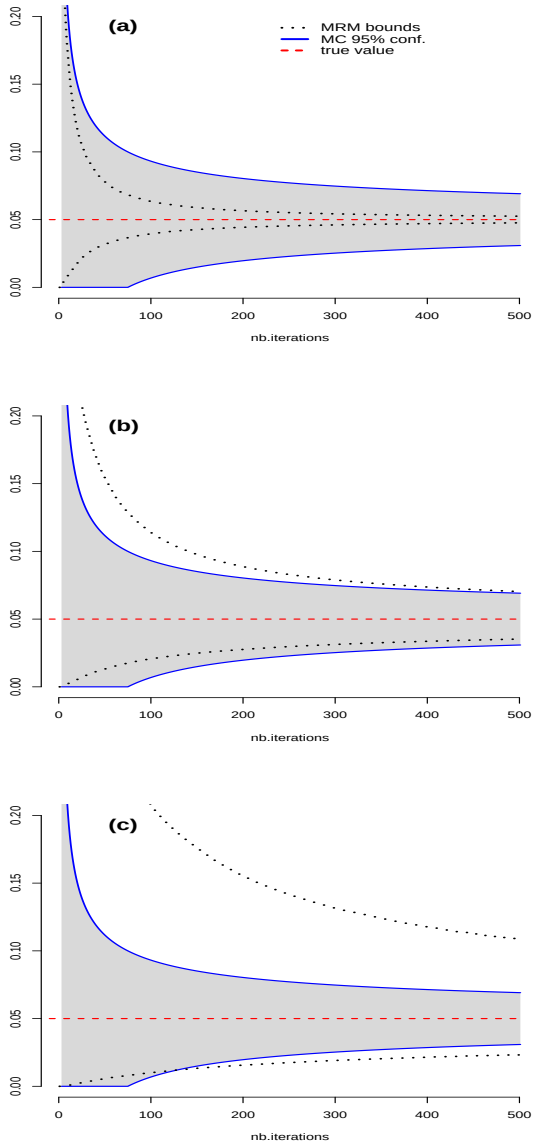

Fig. 3 MRM bounds and MC 95% confidence areas for $n = 300$ trials of G_d with $d = 2$ **(a)**, $d = 3$ **(b)** or $d = 4$ **(c)**. The results are based on a succession of nested uniform samplings in the non-dominated areas and estimated over 300 parallel computations with $p_f = 0.05$.

Note that $V_\infty(2) = 1/2 = p_f$ as it could be expected, but $V_\infty(d) < p_f$ $\forall d \geq 3$; actually $V_\infty(d) \to 0$ when d increases.

An obvious consequence of this quick behavior study is that, as soon higlighted by Limbourg et al. (2010) and shown in Figure 3, the bounds remain very crude conservative estimates of p_f and cannot be practical by themselves in applied studies when d increases. Therefore, ρ_n cannot be used as a precision criterion strictly speaking, but rather as an indicator of reduction speed in space exploration. Hopefully, the bounds may be used to build statistical estimators of p_f to improve significantly some features of the MC estimators.

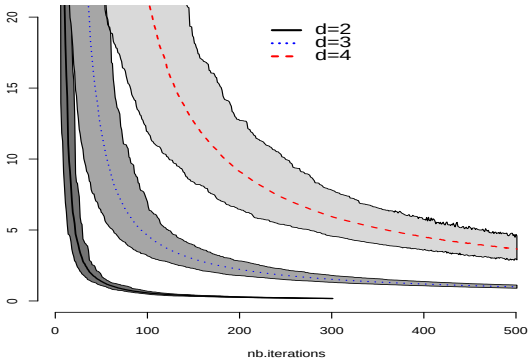

Fig. 4 Plots of convergence criterion ρ_n (median and 95% confidence areas estimated over 300 parallel computations).

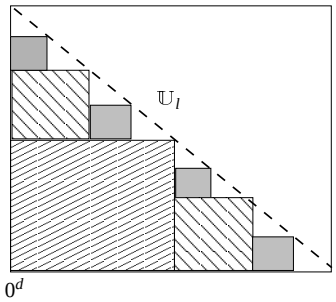

Fig. 5 Fractalized (dichotomic) covering of a half-cube in two dimensions.

Following Limbourg et al. (2010), it could be tempting to consider crude estimators based on an average of these bounds. But since the bounds may not be consistent estimators of p_f, the self consistency of such estimators seems questionable when d is higher than 2. Therefore a cautious elicitation of these estimators is required, as well as the call to probabilistic arguments to study their convergence.

3 A maximum-likelihood estimator of p_f

This section is dedicated to a first approach of the statistical problem of estimating p_f, assuming $\boldsymbol{\nu}_1, \ldots, \boldsymbol{\nu}_{n+1}$ are successively uniformly sampled in the nested non-dominated spaces $\mathbb{U}_0, \ldots, U_n$. The results presented here are useful in three senses. Firstly, the estimation procedure does not need to be calibrated. Secondly, the results highlight asymptotic Monte Carlo acceleration, conservatism and robustness properties in estimation which testify to the interest of taking account of code monotony in any computational framework (without specially parsimonious requirements). Thirdly, the gain in variance with respect to a MC estimator appears as a new objective to overtake, for instance via a generalized importance sampling approach considered in the following section.

When at step k the sampling is uniform, the occurence of a nonzero signature $\xi_{\boldsymbol{\nu}_k}$ follows a Bernoulli distribution $\mathscr{B}(\gamma_k)$, conditional to $\mathscr{F}_{k-1}$, with

$$\begin{aligned}\gamma_k &= P\left(\tilde{G}(\boldsymbol{\nu}) \leq 0 | \boldsymbol{\nu} \in \mathbb{U}_{k-1}\right), \\ &= \frac{P\left(\tilde{G}(\boldsymbol{\nu}) \leq 0\right) - P\left(\tilde{G}(\boldsymbol{\nu}) \leq 0 | \boldsymbol{\nu} \in \mathbb{U}_{k-1}^-\right) P\left(\boldsymbol{\nu} \in \mathbb{U}_{k-1}^-\right)}{P\left(\boldsymbol{\nu} \in \mathbb{U}_{k-1}\right)}\end{aligned}$$

from Bayes' formula, hence

$$\gamma_k = \frac{p_f - p_{k-1}^-}{p_{k-1}^+ - p_{k-1}^-}. \tag{3}$$

After n steps, all information about p_f is thus brought by the dependent-data likelihood $L_n(p_f) = L_n(p_f | \boldsymbol{\nu}_1, \ldots, \boldsymbol{\nu}_n)$ defined by the product of these conditional Bernoulli densities:

$$L_n(p_f) = \prod_{k=1}^{n} \left(\frac{p_f - p_{k-1}^-}{p_{k-1}^+ - p_{k-1}^-}\right)^{\xi_{\boldsymbol{\nu}_k}} \left(\frac{p_{k-1}^+ - p_f}{p_{k-1}^+ - p_{k-1}^-}\right)^{1-\xi_{\boldsymbol{\nu}_k}},$$

the maximum estimator $\hat{p_f}_n$ of which is considered in next proposition.

Proposition 1 *Denote $\ell_n(p_f) = \log L_n(p_f)$. There exists a unique solution $\hat{p_f}_n$ in (p_{n-1}^-, p_{n-1}^+) of the likelihood equation $\ell_n'(p_f) = \sum_{k=1}^n \tilde{\omega}_k\left(p_f\right)\left(p_k - p_f\right) = 0$, semi-explicitely defined by*

$$\hat{p_f}_n = \frac{\sum_{k=1}^{n} \tilde{\omega}_k\left(\hat{p_f}_n\right) p_k}{\sum_{k=1}^{n} \tilde{\omega}_k\left(\hat{p_f}_n\right)},$$

$$\text{where} \quad p_k = p_{k-1}^- + \left(p_{k-1}^+ - p_{k-1}^-\right)\xi_{\boldsymbol{\nu}_k} \tag{4}$$

$$\text{and} \quad \tilde{\omega}_k(x) = \left[\left(x - p_{k-1}^-\right)\left(p_{k-1}^+ - x\right)\right]^{-1}.$$

Numerically a simple Newton-Raphson routine can efficiently do the job. The existence of the MLE $\hat{p_f}_n$ is however restricted to cases where p_f cannot be reached by at least one of the two bounds (p_{n-1}^-, p_{n+1}^-) for any finite n. This is true when a non-empty subset of $\mathbb{U}_l$ is smooth, namely when $\mathbb{U}_l$ is not the surface of union of hyperrectangles sharing the same orthogonal basis, which seems the most likely in practice. Condition (i) in Theorem 1 formalizes this smoothness condition.

Although we are in a data-dependent context, the asymptotic results about $\hat{p_f}_n$ presented in this theorem are classic in the sense that the Cramer-Rao bound for the variance is still given by the inverse of the Fisher information, and is asymptotically reached by the MLE. They are technically based on the martingality of the score process $\{\ell_n'(p_f)\}$.

Theorem 1 *Denote $V_n^{MC}(p_f)$ the variance of a standard MC $n-$estimator and $J_n(p_f) = \sum_{k=1}^{n} E_{\mathscr{F}_k}[\tilde{\omega}_k(p_f)]$ the Fisher information. Assuming that*

(i) $\mathbb{U}_l$ is $\mathscr{C}^1$ in a non-empty measurable subset of $\mathbb{U}$,

then

$$V_n^{MLE} \;=\; J_n^{-1}(p_f) \leq V_n^{MC}(p_f)\left(\frac{2p_f}{1-p_f}\right)\left(\rho_{n-1}^{med}\right)^2.$$

Furthermore, under the supplementary assumptions:

(ii) $\exists\, \delta \in [0,1)$, *with* $\delta = 0$ *only if* $\rho_n^{med} \to 0$, *such that*
$\sum_{k=1}^{n}\left(\tilde{\omega}_k(p_f) - E_{\mathscr{F}_k}\left[\tilde{\omega}_k(p_f)\right]\right) = o\left(n^{1-\delta}\left(\rho_{n-1}^{med}\right)^{-2}\right)$,

(iii) $(p_n^+ - p_f)/(p_f - p_n^-) \xrightarrow[n\to\infty]{I\!P} 1$,

then $\hat{p_f}_n \xrightarrow{I\!P} p_f$

and $\dfrac{1}{\sqrt{V_n^{MLE}}}\left(\hat{p_f}_n - p_f\right) \xrightarrow{\mathscr{L}} \mathscr{N}(0,1)$.

Condition (ii) implies that, when the number of steps comes to infinity, the weights end to vary in a negligible proportion in regards of all possible trajectories of successive uniform nested samplings. Although it is difficult to check, this behavior seems rather reasonable in practice because trajectories mainly vary at the first steps of the algorithm, when the non-dominated space is still large. A close idea, expressed in Condition (iii), is requiring that asymptotically the reduction speeds of intervals $[p_n^-, p_f]$ and $[p_f, p_n^+]$ become similar.

In corollary of the previous theorem, note that asymptotically the coefficient of variation (CV) is such that

$$\mathrm{CV}\left[\hat{p_f}_n\right] \leq \frac{p_f \rho_{n-1}^{\mathrm{med}}}{\mathrm{E}_{\mathscr{F}_n}\left[p_{n-1}^-\right]}\sqrt{\frac{2}{n}}$$

which remains finite at any n when $p_f \to 0^+$. The MLE is thus asymptotically robust.

The variance reduction with respect to $V_n^{MC}(p_f)$ shown in the previous theorem is practical to demonstrate this asymptotic robustness, but remains however submitted to the behavior of $\rho_{n-1}^{\mathrm{med}}$, which is bounded and goes to 0 when $d = 1$, but can stay at high values when d increases (cf. Figure 4). Therefore it must be noticed that from Jensen'inequality,

$$V_n^{MLE} \leq \left(\sum_{k=1}^{n} \mathrm{E}^{-}1\left[\tilde{\omega}_k(p_f)\right]\right)^{-1} = \frac{p_f(1-p_f)}{\sum_{k=1}^{n}(1-c_{k-1})^{-1}} \quad (5)$$

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

$$c_k = \mathrm{E}_{\mathscr{F}_k}\left[\frac{p_k^-}{p_f} + \frac{1-p_k^+}{1-p_f} - \frac{p_k^-(1-p_k^+)}{p_f(1-p_f)}\right] \quad (6)$$

which increases from 0 to 1 when $k \to \infty$.

Thus $V_n^{MLE} \leq \alpha_n(p_f) V_n^{MC}(p_f)$ with

$$\alpha_n(p_f) = \frac{n}{\sum_{k=1}^{n}(1-c_{k-1})^{-1}} \;<\; 1.$$

Proposition 2 *Under the assumptions of Theorem 1, and assuming in addition*

(iv) $\delta > 1/2$ *in Assumption* (ii),
(v) $\nexists\, n \geq 1$ *such that*

$$p_f = (2n)^{-1}\sum_{k=1}^{n}\tilde{\omega}_k(p_f)(p_{k-1}^- + p_{k-1}^-)/\sum_{k=1}^{n}\tilde{\omega}_k(p_f),$$

then

$$\frac{\hat{J}_n^{5/2}(p_f)}{|\hat{J}_n'(p_f)|}\left(\hat{J}_n^{-1}(\hat{p_f}_n) - V_n^{MLE}\right) \xrightarrow{\mathscr{L}} \mathscr{N}(0,1).$$

where $\hat{J}_n(p_f) = \sum_{k=1}^{n}\tilde{\omega}_k(p_f)$.

As a consequence, an asymptotic confidence interval for p_f can be established: denoting u_α the standard normal $\alpha-$order percentile,

$$\lim_{n\to\infty} P\left(\hat{p_f}_n - \frac{u_{1-\alpha/2}}{\sqrt{\hat{J}_n(\hat{p_f}_n)}} \leq p_f \leq \hat{p_f}_n + \frac{u_{1-\alpha/2}}{\sqrt{\hat{J}_n(\hat{p_f}_n)}}\right) = 1-\alpha.$$

Using the recent results by Furrer (2002), strong consistency needs supplementary hypotheses such that $\ell_n''(p_f)/J_n(p_f) \to 1$ almost surely, and checking the Lipschitzian nature of the mapping $\ell_n''(p_f)$: there should exist a sequence $\{\lambda_n\} > 0$ almost surely finite such that, $\forall\ \tilde{p_f}$ in a neighborhood of p_f included in (p_{n-1}^-, p_{n-1}^+),

$$\left|\sum_{k=1}^{n}\left(p_f\tilde{\omega}_k^2(p_f) - \tilde{p_f}\tilde{\omega}_k^2(\tilde{p_f})\right)\right| \leq \lambda_n\left|p_f - \tilde{p_f}\right|.$$

When p_f is low, the MLE $\hat{p_f}_n$ is a conservative estimator of p_f since its bias remains strictly positive, as exemplified on Figure 7. This behavior is due to the fact that the highest weights favor local estimators $p_k = p_k^+$ when approaching the failure surface $\mathbb{U}_l$ ($\xi_{\boldsymbol{\nu}_k} = 1$ in (4). On the exemple considered in this last figure (as well as in other experiments not shown 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. However, note that from a strictly statistical criterion (root mean square error), all MLE in dimensions 2 to 4 for the generic example end to improve the standard Monte Carlo estimator (Figure 6, *left*).

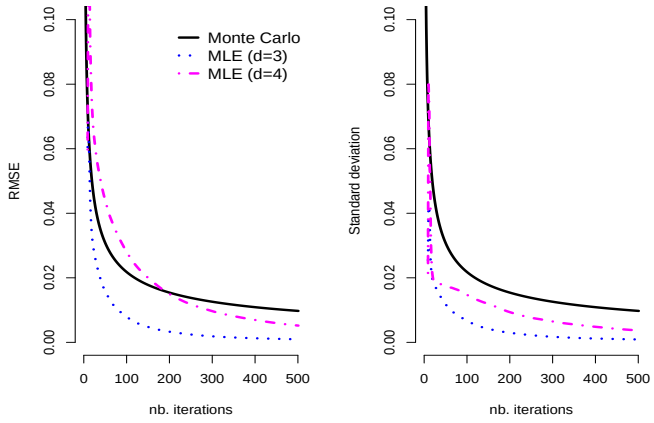

Fig. 6 Root mean square error (RMSE; *left*) and standard deviation (*right*) of the standard Monte Carlo estimator and $\hat{p_{f}}_n$ for $d = 3$ and $d = 4$, for $p_f = 0.05$. Empirical estimations are made over 300 parallel MRM runs.

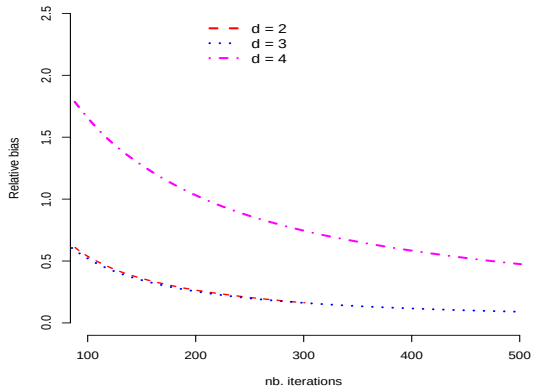

Fig. 7 Relative bias of the MLE $\hat{p}_n$ for the dimensions $d \in \{2,3,4\}$.

Controlling and correcting the bias could be the subject of a research based on bootstrap experiments, using a monotonous estimation of the failure surface $\mathbb{U}_l$ calibrated from the last obtained non-dominated points in $\mathbb{U}$. However, naturally unbiased estimators with non-asymptotic properties appear more relevant in our costly-computational framework. Rather than the usual Monte Carlo variance, getting variances of such unbiased estimators that are below the asymptotic variance of the MLE should become an aim. Because the latter reaches the Cramer-Rao bound, this needs to consider other forms of sampling than (naive) uniform ones, in the succession of nested non-dominated spaces, to speed up the volume removing. Next section provides general results to handle and select these samplings.

4 Weighted importance sampling estimators of p_f

4.1 Definition

Now denote $f_{k-1}(\boldsymbol{\nu})$ any importance sampling pdf defined on $\mathbb{U}_{k-1}$, from which the new point of the DoE, $\boldsymbol{\nu}_k$, is sampled at iteration k. Denote

$$\tilde{p}_k = p_{k-1}^- + \frac{\xi_{\boldsymbol{\nu}_k}}{f_{k-1}(\boldsymbol{\nu}_k)},$$

thus $\tilde{p}_k$ can be seen as a generalization of estimator p_k defined in (4). We obviously assume that $\mathrm{Supp}(f_{k-1}) = \mathbb{U}_{k-1}$ such that $\tilde{p}_k$ is always well defined. Notice that

$$\tilde{p}_k \in \left[p_{k-1}^-, p_{k-1}^+\right] \Leftrightarrow f_{k-1}(\boldsymbol{\nu}_k) \geq \left(p_{k-1}^+ - p_{k-1}^-\right)^{-1} \tag{7}$$

Lemma 2 *Estimator $\tilde{p}_k$ is unbiased.*

Consequently, the weighted importance sampling estimator (WISE)

$$\breve{p}_{f\,n} = \frac{1}{n}\sum_{k=1}^{n} \omega_k \tilde{p}_k, \tag{8}$$

where the ω_k are deterministic weights in $[0,n]$, the sum of which being equal to n, is unbiased too. Its variance is given in next proposition.

Proposition 3 *The variance of $\breve{p}_{f\,n}$ is*

$$V_n^{WISE} = \sum_{k=1}^{n} \frac{\omega_k^2}{n^2} E\left[\int_{\mathbb{U}_{k-1}} \frac{\mathbb{1}_{\{\tilde{G}(\boldsymbol{\nu})\leq 0\}}}{f_{k-1}(\boldsymbol{\nu})}\, d\boldsymbol{\nu} - (p_f - p_{k-1}^-)^2\right],$$

expectations being defined with respect to $\mathscr{F}_{k-1}$. An alternative writing is $V_n^{WISE} = V_n^U(p_f) + V_n^F(p_f)$ where

$$V_n^U(p_f) = V_n^{MC}(p_f) \sum_{k=1}^{n} \frac{\omega_k^2}{n}(1 - c_{k-1}),$$

$$V_n^F(p_f) = \sum_{k=1}^{n} \frac{\omega_k^2}{n^2} E\left[\int_{\mathbb{U}_{k-1}} \frac{\mathbb{1}_{\{\tilde{G}(\boldsymbol{\nu})\leq 0\}}}{f_{k-1}(\boldsymbol{\nu})}\, d\boldsymbol{\nu} - \frac{(p_f - p_{k-1}^-)}{(p_{n-1}^+ - p_{n-1}^-)^{-1}}\right]$$

with $V_n^F(p_f) = 0$ if all the f_{k-1} are chosen uniform, the c_k's being defined in (6).

4.2 General properties

The features of this variance deserves some comments.

(a) In the most naive case of sampling, namely when $\omega_k = 1$ and f_{k-1} is uniform on $\mathbb{U}_{k-1}$, then a MC approach remains always beaten since

$$V_n^{\mathrm{U}}(p_f) = V_n^{MC}(p_f)\left(1 - \frac{1}{n}\sum_{k=1}^{n} c_{k-1}\right) < V_n^{MC}(p_f).$$

(b) Surprisingly, V_n^{WISE} is the weighted sum of the variances of estimators $\tilde{p}_k$, $k \leq n$ (see proof), so that they are decorrelated and the weighted sum of unbiased estimators of $\mathrm{Var}[\tilde{p}_k]$ provides an unbiased estimator of V_n^{WISE}. For instance, for $k < n$, defining $\hat{V}_k = (\tilde{p}_k - p_{k-1}^-)(\tilde{p}_k - \tilde{p}_{k+1})$, one has $\mathrm{E}\left[\hat{V}_k\right]$ equal to

$$\begin{aligned}
&\mathrm{E}\left[\mathrm{E}_{f_{k-1}}\left[\frac{\xi_{\boldsymbol{\nu}}}{f_{k-1}(\boldsymbol{\nu})}\left(\frac{\xi_{\boldsymbol{\nu}}}{f_{k-1}(\boldsymbol{\nu})} + p_{k-1}^- - \mathrm{E}\left[\tilde{p}_{n+1}|\mathscr{F}_k\right]\right)\right]\right], \\
&= \mathrm{E}\left[\int_{\mathbb{U}_{k-1}} \frac{\xi_{\boldsymbol{\nu}}}{f_{k-1}(\boldsymbol{\nu})}\, d\boldsymbol{\nu} - (p_f - p_{k-1}^-)\int_{\mathbb{U}_{k-1}} \xi_{\boldsymbol{\nu}}\, d\boldsymbol{\nu}\right], \\
&= \mathrm{E}\left[\int_{\mathbb{U}_{k-1}} \frac{\xi_{\boldsymbol{\nu}}}{f_{k-1}(\boldsymbol{\nu})}\, d\boldsymbol{\nu} - (p_f - p_{k-1}^-)^2\right], \\
&= \mathrm{Var}\left[\tilde{p}_k\right].
\end{aligned}$$

(c) Assuming the $\{f_{k-1}\}$ are known, the weights $\{\omega_k\}$ can be theoretically calibrated such that the variance be minimized, as explained in Corollary 1.

Corollary 1 *Denote $d_0 = 0$ and $d_k = c_k - E[b_k(f_k)]/(p_f(1-p_f))$ $\forall k > 1$ where*

$$b_k(f_k) = \int_{\mathbb{U}_k} \frac{\mathbb{1}_{\{\tilde{G}(\boldsymbol{\nu})\leq 0\}}}{f_k(\boldsymbol{\nu})}\, d\boldsymbol{\nu} - \frac{(p_f - p_k^-)}{(p_n^+ - p_n^-)^{-1}}.$$

The solution of the optimization problem

$$(\omega_1^*, \ldots, \omega_n^*) = \arg\min V_n^{WISE},$$

under the constraint $\sum_{k=1}^n \omega_k = n$, is

$$\omega_k^* = \frac{n}{(1-d_{k-1})\sum_{j=1}^n (1-d_{j-1})^{-1}}. \tag{9}$$

Consequently, the optimized variance is

$$V_n^{*WISE} = V_n^{MC}(p_f)\frac{n}{\sum_{k=1}^n (1-d_{k-1})^{-1}}.$$

(d) When all f_{k-1} are chosen uniform, then $d_k = c_k$ and

$$V_n^{*WISE} = \left(\sum_{k=1}^n \mathrm{E}^{-1}\left[\tilde{\omega}_k^{-1}(p_f)\right]\right)^{-1} \geq V_n^{MLE}$$

from (5). This result was expectable since V_n^{MLE} is the Cramer-Rao bound for this kind of sampling. This testifies from the need of eliciting carefully the importance pdf to diminish the variance.

(e) Eliciting importance pdf $\{f_{k-1}\}$ such that $b_{k-1}(f_{k-1}) < 0$ for all k ensures the non-asymptotic robustness of the estimator $\breve{p}_{fn}$, especially when the weights are optimized. Next proposition details this property. Note that this condition is automatically checked if f_{k-1} is elicited such that (7) is true for all $\boldsymbol{\nu}_k \in \mathbb{U}_{k-1}$.

Proposition 4 *Assume that for any $k \geq 1$,*

$$\int_{\mathbb{U}_{k-1}} \frac{\mathbb{1}_{\{\tilde{G}(\boldsymbol{\nu})\leq 0\}}}{f_{k-1}(\boldsymbol{\nu})}\, d\boldsymbol{\nu} \leq (p_f - p_{k-1}^-)(p_{n-1}^+ - p_{n-1}^-). \tag{10}$$

Then $\exists\, C > 0$ such that

$$CV\left[\breve{p}_{fn}\right] \leq \frac{C}{n}\sqrt{\sum_{k=1}^n \omega_k^2} \tag{11}$$

and $\breve{p}_{fn}$ is also a robust estimator of p_f.

(f) As usual in importance sampling approaches in structural reliability (Rubino and Tuffin 2009), a particular choice of the importance density (in our case, of successive importance densities) can reduce the variance of the estimator to 0: see Proposition 5. This density is obviously not elicitable in practice since it requires to know the failure surface $\mathbb{U}_l$ and the probability p_f that we are precisely looking for. But this result illustrates the potentiality of strong variance reduction induced by importance sampling approaches carrying out approximations of $\mathbb{U}_l$.

Proposition 5 *A null variance $V_n^{WISE} = 0$ can be reached if and only if, at each step n, $\exists k_{n-1}(\boldsymbol{\nu})$ a density on $\mathbb{U}_n^+ = \mathbb{U}^+ \cap \mathbb{U}_{n-1}$ such that*

$$f_{n-1}(\boldsymbol{\nu}) = \begin{cases} (p_f - p_{n-1}^-)^{-1} & \text{if } \tilde{G}(\boldsymbol{\nu}) \leq 0, \\ k_{n-1}(\boldsymbol{\nu}) & \text{else} \end{cases} \tag{12}$$

and, $\forall \boldsymbol{\nu}_l \in \mathbb{U}_l$,

$$\lim_{\varepsilon\to 0}\int_{B_{\boldsymbol{\nu}_l}^{\varepsilon}} k_{n-1}(\boldsymbol{\nu})\, d\boldsymbol{\nu} = \frac{Vol\left(\mathbb{U}_{\boldsymbol{\nu}_l}^- \cap \mathbb{U}_{n-1}\right)}{p_f - p_{n-1}^-}$$

where $B_{\boldsymbol{\nu}_l}^{\varepsilon} = \{\boldsymbol{\nu}_l + u\varepsilon,\ u \in [0,1]^d\}$ is the open half-ball of radius $\varepsilon \geq 0$ centered on $\boldsymbol{\nu}_l$ containing safety points only.

More realistically, constraining $b_{k-1}(f_{k-1}) \leq 0$ can be achieved giving more weight to a failure-dominated area with maximal volume, namely f_{k-1} should sample close to $\mathbb{U}_l$. Tthis intuitive reasoning was the rationale for Limbourg et al. (2010) to propose an empirical directional strategy at each MRM step. The benefit of such a strategy is illlustrated in next toy example.

Example 3 Consider a unidimensional problem where $\mathbb{U}_l = \{\boldsymbol{\nu}_S\} = \{p_f\} \in [0,1]$. After $k-1$ iterations, denote $\boldsymbol{\nu}_{k-1}^- = p_{k-1}^-$ and $\boldsymbol{\nu}_{k-1}^+ = p_{k-1}^+$. Assuming to known a priori a perfect estimation of $\mathbb{U}_l$ (ie., with negligible error), let choose $f_{k-1}(\boldsymbol{\nu})$ as the truncated Cauchy distribution on $[\boldsymbol{\nu}_{k-1}^-, \boldsymbol{\nu}_{k-1}^+]$ with mode $\boldsymbol{\nu}_S$ and scale γ. Then, after basic algebra,

$$b_{k-1}(f_{k-1}) = (\boldsymbol{\nu}_S - \boldsymbol{\nu}_{k-1}^-)\left[\Delta\frac{(\gamma + \boldsymbol{\nu}_S - \boldsymbol{\nu}_{k-1}^-)^2/3}{\boldsymbol{\nu}_{k-1}^+ - \boldsymbol{\nu}_{k-1}^-} - 1\right]$$

with $\Delta = \arctan((\boldsymbol{\nu}^+_{k-1} - \boldsymbol{\nu}_S)/\gamma) + \arctan((\boldsymbol{\nu}_S - \boldsymbol{\nu}^-_{k-1})/\gamma) \leq \pi$ so that, assuming the mild condition $\boldsymbol{\nu}_S - \boldsymbol{\nu}^-_{k-1} < \sqrt{3/\pi} \simeq 0.977$, one must select $\gamma < \pi^{-1} - (\boldsymbol{\nu}_S - \boldsymbol{\nu}^-_{k-1})^2/3$ to ensure $b_{k-1}(f_{k-1}) < 0$, and a decreasing γ (ie., a more concentrated importance distribution around $\boldsymbol{\nu}_S$) increases $|b_{k-1}(f_{k-1})|$.

Then Remarks (**e**) and (**f**) show that a realistic sequence of good importance pdf $\{f_k\}$, cleverer than simple uniforms, must be elicited based on a sequence of *surrogates* of the failure surface $\mathbb{U}_l$. Next subsection is dedicated to a heuristic (work-in-progress) approach of the elicitation of $\{f_k\}$ in practice.

4.3 Heuristic elicitation of $\{f_k\}$ in practice

Since an arbitrary number of points can be independently sampled in both dominated areas at each step, $\mathbb{U}_l$ can be easily interpreted as the decision frontier of a supervised classification binary problem, without horseriding of classes (perfectly separable). A nonparametric estimation $\hat{\mathbb{U}}_{k,l}$ of the frontier $\mathbb{U}_l$ can be done using Multi-Layer Perceptron (MLP) neural networks or Support Vector Machines (SVM), as recommended by Hurtado (2004) in the field of structural reliability analysis. Li et al. (2006) emphasized that such tools are flexible, can estimate a frontier on the basis of a few samples, and can overcome the curse of dimensionality. To be realistic and coherent with the features of G, two constraints must be applied on the estimation of $\hat{\mathbb{U}}_{k,l}$:

- the volume of $\hat{\mathbb{U}}^-_k = \{\boldsymbol{\nu} \in \mathbb{U},\ \exists \boldsymbol{\nu}_l \in \hat{\mathbb{U}}_{k,l},\ \boldsymbol{\nu} \preceq \boldsymbol{\nu}_l\}$ must be equal to the current estimator $\check{p}_{f_k}$ of p_f;
- $\hat{\mathbb{U}}_{k,l}$ must be a decreasing function of the coordinates.

Once the classifier $\hat{\mathbb{U}}_{k,l}$ has been obtained, location and scale parameters for f_{k-1} must be elicited (assuming a form for f_{k-1} has been chosen) such that the constraints

$$\tilde{p}_k \in \left[p^-_{k-1}, p^+_{k+1}\right] \tag{13}$$

$$\check{p}_{f_k} \in \left[p^-_{k-1}, p^+_{k+1}\right] \text{ (if possible)}, \tag{14}$$

$$\int_{\hat{\mathbb{U}}^-_k} \frac{1}{f_k(\boldsymbol{\nu})}\, d\boldsymbol{\nu} \leq \left(\check{p}_{f_k} - p^-_{k-1}\right)\left(p^+_n - p^-_{n-1}\right) \tag{15}$$

are respected. Typically, choosing a small scale parameter (or a small importance variance) will lead to satisfy the boundary constraints (13) and (14), as exemplified in next proposition when using truncated multinormal importance pdf.

Proposition 6 *For $d \geq 2$, let $\mu_{n,d}$ be the unique positive solution of equation $D(x) = x^d + dx^{d-2} = (p^+_n - p^-_n)/(2\pi)^{d/2}$. Note that $\mu_{n,d}$ strictly decrease with n, following the decreasing of $p^+_n - p^-_n$). For $q > d-2$, let $\{\eta_n\}$ be a sequence of values in $[0,1]$ such that, for $d > 2$,*

$$\eta_n = o\left(n^{-q/(d-2)} d^{-1/(d-2)} (2\pi)^{-d/2(d-2)}\right) \tag{16}$$

and define $\sigma_n = \mu_{n,d}\eta_n$. Finally, let f_n be a multinormal distribution with mean $\boldsymbol{\nu}_{n,0}$ and variance $\sigma_n^2 I_d$, truncated on $\mathbb{U}_n$. Then

$$p^-_n \leq \tilde{p}_{n+1} \leq p^+_n + o\left(n^{-q}\right).$$

The quasi-determinism implied by small importance variances has two major consequences.

(i) A sampled $\boldsymbol{\nu} \sim f_{k-1}$ remains close to the location parameter $\boldsymbol{\nu}_{k-1,0}$, which takes the sense of the point of $\mathbb{U}_{k-1}$ such that, given the classifier $\hat{\mathbb{U}}_{k,l}$, the maximal dominated volume can be removed at step k; based on the equality $1 - c_k = \mathrm{E}[(p_f - p^-_k)(p^+_k - p_f)]$, a simplified criterion to minimize in $\boldsymbol{\nu}_{k-1,0}$ could be

$$C_{k-1}(\boldsymbol{\nu}) = \left(\check{p}_{f_{k-1}} - p^-_k(\boldsymbol{\nu})\right)\left(p^+_k(\boldsymbol{\nu}) - \check{p}_{f_{k-1}}\right) \tag{17}$$

where $p^{-+}_k(\boldsymbol{\nu})$ is the bound implied by the position of $\boldsymbol{\nu}$ with respect to $\hat{\mathbb{U}}_{k,l}$. See Example 4 for a crude illustration of the potential benefits of this approach.

(ii) The sequence $\{p^-_k, p^+_l\}$ becomes quasi-deterministic too, in the sense that $\mathrm{Var}[p^-_n]$ and $\mathrm{Var}[p^+_n]$ can decrease toward 0. As seen in next subsection, such a behavior appears as a sufficient condition for the distribution of $\check{p}_{f_n}$ to be asymptotically normal.

Example 4 Treating the generic example in dimension $d = 2$, a simple MLP with 3 hidden layers was calibrated at each iteration. It was found decreasingly monotonous on nearly 90% of the iterations, and its prediction error was rarely found over 15%. At iteration n, 10^3 points were uniformly sampled in the non-dominated space. The mode of a truncated gaussian importance pdf was selected among these points by minimizing the criterion (17). Its standard deviation was chosen as $\sigma_n = n^{-2}(2\pi)^{-2}\mu_{n,2}$. Despite the crudity of this elicitation, the decreasing of the width between the bounds was found significantly improved with respect to the naive approach based on uniform sampling (Figure 8), as well as the precision of the estimators (Figure 9).

4.4 Asymptotic behavior

Because all estimators $\tilde{p}_k$ are dependent, finding the non-asymptotic distribution of $\check{p}_{f_n}$ seems rather difficult. The asymptotic behavior of $\check{p}_{f_n}$, which is also needed to determine the behavior of variance and weights empirical estimators, can however be studied using tools from martingale theory similar to those that have been used to explore the asymptotic behavior of the MLE. This requires an autoregressive writing of $\check{p}_{f_n}$, which is simple to get. Indeed, denoting

$$\beta_n = \frac{\omega_n^{*[n]}}{n}$$

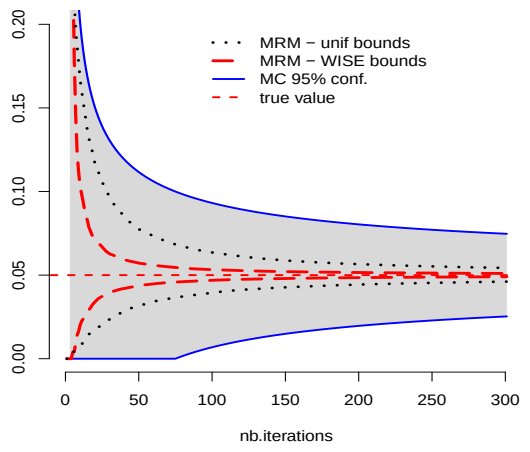

Fig. 8 MRM bounds (uniform / importance samplings) and 95% Monte Carlo confidence bounds ($d = 2$). Results are averaged over 300 repeated simulations.

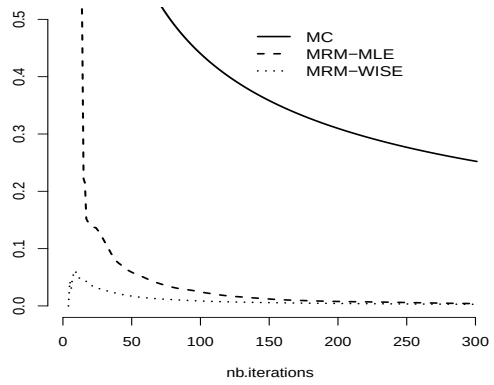

Fig. 9 Coefficients of variation for the MC, MLE and WISE estimators ($d = 2$).

where $\omega_k^{*[n]}$ is the k^{th} optimized weight associated to the $n-$estimator $\check{p}_{f_n}$, then

$$\check{p}_{f_{n+1}} = (1-\beta_{n+1})\,\check{p}_{f_n} + \beta_{n+1}\tilde{p}_{n+1}. \tag{18}$$

Then the following results are based on the martingality of the process $\{Z_n\}$ defined by $Z_n = \lambda_n(\check{p}_{f_n} - p_f)$ where $\lambda_n = \prod_{k=1}^{n}(1-\beta_k)^{-1}$.

Theorem 2 *Under the following assumptions:*

(i) $p_n^- - E[p_n^-] \xrightarrow{\mathbb{P}} 0$, $p_n^+ - E[p_n^+] \xrightarrow{\mathbb{P}} 0$,

(ii) $Cov[p_n^-, p_n^+] \xrightarrow{\mathbb{P}} 0$,

(iii) $b_n(f_n) - E[b_n(f_n)] \xrightarrow{\mathbb{P}} 0$,

then $\check{p}_{f_n} \xrightarrow{s.} p_f$

and $$\frac{(\check{p}_{f_n} - p_f)}{\sqrt{V_n^{*WISE}}} \xrightarrow{\mathscr{L}} \mathcal{N}(0,1).$$

From Chebychev's inequality, Condition (i) in Theorem 2 is checked if $\mathrm{Var}[p_n^-]$ and $\mathrm{Var}[p_n^+]$ can decrease toward 0, which may be ensured by a quasi-deterministic choice of the trial $\boldsymbol{\nu}_k$ at iteration k, as explained before.

To go more in details, consider only the behavior of p_n^-. Notice that

$$p_n^- = p_{n-1}^- + \xi_{\boldsymbol{\nu}_n}\mathrm{Vol}_{n-1}^-(\boldsymbol{\nu}_n)$$

where $\mathrm{Vol}_{n-1}^-(\boldsymbol{\nu})$ is the element of volume added to $\mathbb{U}_{n-1}^-$, namely the volume of the set of all points of $\mathbb{U}_n$ located under $\boldsymbol{\nu}$. Therefore one has

$$\begin{aligned}\mathrm{Var}[p_n^-] = \mathrm{Var}[p_{n-1}^-] &+ 2\mathrm{Cov}\left[p_{n-1}^-, \mathrm{E}_{f_{n-1}}\left[\xi_{\boldsymbol{\nu}}\mathrm{Vol}_{n-1}^-(\boldsymbol{\nu})\right]\right] \\ &+ V\left[\xi_{\boldsymbol{\nu}}\mathrm{Vol}_{n-1}^-(\boldsymbol{\nu})\right]\end{aligned}$$

with

$$\begin{aligned}V\left[\xi_{\boldsymbol{\nu}}\mathrm{Vol}_{n-1}^-(\boldsymbol{\nu})\right] = \mathrm{Var}&\left[\mathrm{E}_{f_{n-1}}\left[\xi_{\boldsymbol{\nu}}\mathrm{Vol}_{n-1}^-(\boldsymbol{\nu})\right]\right] \\ &+ \mathrm{E}\left[\mathrm{Var}_{f_{n-1}}\left[\xi_{\boldsymbol{\nu}}\mathrm{Vol}_{n-1}^-(\boldsymbol{\nu})\right]\right]\end{aligned}$$

Formally, $\mathrm{Var}[p_n^-]$ can decrease toward 0 when the sampling is quasi-deterministic in the sense discussed in the next items.

1. Denote $\delta_{\boldsymbol{\nu}}$ the Dirac measure in $\boldsymbol{\nu}$. Assuming the sequence $\{f_n\}_n$ can be elicited such that $f_n(\boldsymbol{\nu}) - \delta_{\boldsymbol{\nu}_{0,n}}$ converges weakly to 0 allows for $V[\xi_{\boldsymbol{\nu}}\mathrm{Vol}_{n-1}^-(\boldsymbol{\nu})]$ to converge toward 0 in the same sense, using Prohorov's theorem. It simply requires the tightness of the set of measures $\{f_n\}_n$ and the boundness of their two first moments, which are both ensured selecting bounded $\{f_n\}_n$ on $[0,1]^d$.
2. Furthermore, p_{n-1}^- is obviously an increasing function of the $\{\boldsymbol{\nu}_i\}_{1\le i\le n-1}$. Assuming that $\boldsymbol{\nu}_n$ is selected such that the additive dominated volume is maximized at iteration n, then $\mathrm{E}_{f_{n-1}}[\xi_{\boldsymbol{\nu}}\mathrm{Vol}_{n-1}^-(\boldsymbol{\nu})]$ is a decreasing function of the $\{\boldsymbol{\nu}_i\}_{1\le i\le n-1}$. Since these two random variables are bounded then, from Schmidt (2003), $\exists N \in \mathbb{N}$, $\forall n \ge N$,
$$\mathrm{Cov}\left[p_{n-1}^-, \mathrm{E}_{f_{n-1}}\left[\xi_{\boldsymbol{\nu}}\mathrm{Vol}_{n-1}^-(\boldsymbol{\nu})\right]\right] \le 0.$$

As a consequence of an importance sampling close to determinism, the sequence of spaces $\{\mathbb{U}_n^-\}$ becomes itself nearly deterministic, which allows for the condition (iii) in Theorem 2 to be checked.

Finally, Condition (ii) in Theorem 2 says that asymptotically, the probability of improving one bound becomes decorrelated on the probability of improving the other. This can be ensured by the symmetry of the volumes of the non-dominated spaces $(\mathbb{U}_n^-, \mathbb{U}_n^+)$, which seems a relatively mild condition in asymptotic terms (ie., when the frontiers of the dominated sets become very close to the failure surface $\mathbb{U}_l$).

Despite those theoretical conditions remain difficult to check in practice, we noticed that even a single uniform

nested sampling can lead to observe variance decreasing for the bounds and asymptotic normality of the WISE estimator. We are thus confident in the fact that those conditions can be relaxed in future studies.

4.5 Controlling the WISE estimators

As an important feature of the estimator, it is possible to bound its convergence speed in function of the decreasing series of criteria $\rho_1 > \ldots > \rho_n$. Practical usefulness of this result is to help to stop the computation when reaching a predetermined estimation error. Next proposition summarizes this result under the form of a non-asymptotic Hoeffding-type concentration inequality, which does not depend on the distribution of the estimator $\check{p_{f_n}}$.

Anti-conservatism can be defined by the fact that at any step n, the estimator $\check{p_{f_n}}$ can underestimate p_f. Obviously the probability of such an event is asymptotically 1/2 because of the asymptotic normal distribution of $\check{p_{f_n}}$. In a non-asymptotic framework, however, it can be formalized as the probability

$$p_{u,n}(\varepsilon) = P\left(\frac{\check{p_{f_n}}}{p_f} \leq 1-\varepsilon\right).$$

that the computed estimator $\check{p_{f_n}}$ is slightly under p_f, the "slightly" term being associated to a given (possibly normative) precision ε. Upper bounds for this probability can then be provided using the martingality arguments used in the asymptotic study. A first result is given in the following proposition.

Proposition 7

$$p_{u,n}(\varepsilon) \leq \exp\left(-2(p\varepsilon)^2\left[\frac{\left(\sum_{k=1}^{n}(1-d_{k-1})^{-1}\right)^2}{\sum_{k=1}^{n}(1-d_{k-1})^2/f_{k-1}^2(\boldsymbol{\nu}_k)}\right]\right)$$

---

5 Discussion

5.1 Main results and mains concerns

In this article, we have formally described the main features of methods taking advantage of monotony for bounding and estimating probabilities p_f of undesirable events, in the common case when such events are defined as a threshold exceedance by the output of a time-consuming black box. They are based on sequential designs of experiments. Though the sequential bounds around p_f seem not being able to be considered as both consistent and conservative estimators of p_f, except in dimensions 1 and 2, a parsimonious estimation of p_f can be led by statistical means and theoretically studied by probabilistic arguments. Two family of estimators have been proposed, the properties of each improving significantly those of standard Monte Carlo estimators.

If the first one allows for simple computation and leads to a new target variance (rather than the Monte Carlo variance) in this monotonous context, it can suffer from too conservative bias. The second family of estimators, based on sequential importance sampling, appears as the most promising result. However, more work is needed to get a fine calibration of importance distributions through a criterion to minimize, and estimating the various quantities involved in the procedure (importance weights, estimator variance). As detailed in § 4.2, future studies should especially focus on building sequential emulations (or *surrogates*) of the limit state surface $\{\mathbf{x} \in \mathbb{D},\ G(\mathbf{x}) = 0\}$ under monotony constraints. Such constraints have began to be studied in the supervised classification area (see Pelckmans et al. 2005 and Lang 2005, among others), which efficiently stand up the curse of dimensionality, but remain to be adapted to our framework.

The framework considered here remains general and let us hope in a wide range of interesting theoretical researches and possible applications. In this regard, our concern was to provide theoretical and applied tools to allow direct implementation, not only in the specific area of structural reliability where these methods have been heuristically proposed. In the following points, we briefly describe other points of technical interest as other research avenues.

5.2 Relaxing hypotheses on input $\mathbf{X}$

A first significant issue is relaxing the hypotheses made on the input vector $\mathbf{X}$. If the stochastic ordering of the distribution function $F_{\mathbf{X}}$ must be known in absence of another clear transform $\Psi_{F_{\mathbf{X}}}$ taking into account the correlations between components of $\mathbf{X}$, Lemma 3 in Appendix A appears somewhat restrictive. One could think, for instance, that some conditions on the copula function induced in the input representation must be checked to ensure that the monotony of transformed code $\tilde{G}$ is preserved (as those obtained in Chen 2009 about normal copulas). Because inputs can be summarizable in many different ways (through correlation matrices, copulas and marginal distributions, hierarchical buildings...), looking for various conditions of monotony is a practical response to applied concerns. A connected work must be done about the relaxation of monotony constraints : one

should explore the conditions for which the methods developed here can be applied with success to computer codes with local monotonous properties.

5.3 Behavior and control of estimators

The difficulty to check asymptotic conditions and the non-asymptotic behavior of estimators can remain tricky issues, which however can be partially solved using control theory. Estimation results could be enhanced focusing on the convergence properties of estimator $\breve{p_{f}}_{n}$. The result given in (19) must be seen as a first step toward methods for controlling the estimation error (and also the convergence speed), in function of precision criteria $\rho_1, \ldots, \rho_n$. As a sum of dependent random variables, one can hope that Stein's convergence, more refined concentration theorems and large deviations results (Chatterjee 2007, Kontorovich and Ramanan 2008) are applicable too, resulting in a better statistical control. Ideally, better concentration results could be obtained if the variance of $\breve{p_{f}}_{n}$ can be incorporated into the convergence bounds, producing Bernstein-type or Bennett-type inequalities as alternatives to the Azuma-Hoeffding inequality (19), similarly to the standard case of the sum of iid random variables.

5.4 Relaxing hypotheses on the failure surface

The lack of consistency of the bounds observed in practice and theoretically could be overcame by supplementary hypotheses made on the regularity of the failure surface. For instance, it seems natural to link the vertexes by additive segments such that the bounds can be redefined as volumes of their convex hulls. This implies that the failure surface does not present, in the uniform space, a very jagged shape. However, since this feature appears somewhat difficult to check in practice, it should be accepted that the interval between the bounds is no more of 100%-confidence.

5.5 Sensitivity studies

For a given $n-$level of granularity, studies of the sensitivity of p_f, its bounds (p_n^-, p_n^+) and its various estimators to modifications of inputs must be, as emphasized by many authors (references within de Rocquigny et al. 2008), crucial tasks in future applied studies.

As a supplementary benefit from monotony, it must be noted that spaces $(\mathbb{U}_n^-, \mathbb{U}_n^+)$ remain dominated whatever the choice made on input distributions. Indeed, they purely reflect the properties of the deterministic function G, independently of the density of each corresponding sample point in $\mathbb{D}$. An immediate and pleasant consequence is the possibility to recompute the bounds without any supplementary call to G. The only difference lies in the differing probability weight associated to those spaces when the uncertainty model changes from $f_{\mathbf{X}}$ to $f_{\mathbf{X}}^{\varepsilon}$: one just needs to recompute the hypervolumic calculations (De Rocquigny, *private communication*)

$$p_n^-(\varepsilon) = \int_{\Psi_{F_{\mathbf{X}}}^{-1}(\mathbb{U}_n^-)} f_{\mathbf{X}}^{\varepsilon}(\mathbf{x})\, d\mathbf{x},$$

$$p_n^+(\varepsilon) = 1 - \int_{\Psi_{F_{\mathbf{X}}}^{-1}(\mathbb{U}_n^+)} f_{\mathbf{X}}^{\varepsilon}(\mathbf{x})\, d\mathbf{x}$$

for instance by a simple Monte Carlo method. In these future studies, we suggest that the progressive bounds could be defined as *robust* while they remain true whatever the fluctuation of $f_{\mathbf{X}}$ in a given variational class.

5.6 Comparing with and taking benefit of other approaches

Last but not least, future theoretical and applied studies will have to be compared with (and possibly take benefit from) approaches developed in other areas. For instance, evolutionary algorithms are currently studied in multi-objective optimization for covering a maximal hypervolume at each step (Beume et al. 2007). In high-dimensional problems, "divide-and-conquer" strategies of designing experiments can benefit from the increasing use of parallel computing (Wilkinson 2008). In this regard, it will be essential to combine the knowledge of computer code monotony with the will to explore primarily the input space along most relevant dimensions.

Acknowledgements The author thanks Etienne de Rocquigny (École Centrale Paris) for giving him the opportunity to work on this subject and his support, and Philip Limbourg (University of Duisburg-Essen) for some fruitful advices about sweepline algorithms. Guillaume Bouchard (Xerox Research Centre Europe) must be thanked for enlightening discussions about the convergence of the bounds. Many thanks to Merlin Keller (ENGREF) for his careful reading of the paper.

References

1. Au, S., Beck, J.L., Estimation of small failure probabilities in high dimensions by subset simulation, Probabilistic Engineering Mechanics, 16(4), 263-277 (2001).
2. Bercu, B., Inégalités exponentielles pour les martingales, Journées ALEA 2008 CIRM, 10-14 March (2008) [*in French*] .
3. Beume, N., Naujoks, B., Emmerich, N., SMS-EMOA: multiobjective selection based on dominated hypervolume, European Journal of Operational Research, 181(3), 1653-1669 (2007).
4. Cadre, B., Stabilité des médianes conditionnelles par rapport à des filtrations discrétisées, C. R. Acad. Sci. Paris, 335, 545-548 (2002) [*in French*].

5. Cannamela, C., Garnier, J., Iooss, B., Controlled stratification for quantile estimation, Ann. Appl. Stat., 2, 1554-1580 (2008).
6. Chatterjee, S., Stein's method for concentration inequalities, Probability theory and related fields, 138, 305-320 (2007).
7. Chan, T.M., A (slightly) faster algorithm for Klee's measure problem, In: Proceedings of the twenty-fourth annual symposium on Computational Geometr, College Park, MD, USA, 94-100 (2008).
8. Chen, G., Monotonicity of dependence concepts: from independent random vector into dependent random vector, Proceedings of the ICRSSE 2009 Conference (2009).
9. Chlebus, B.S., On the Klee's measure problem in small dimensions, In: Proceedings of the 25th Conference on Current Trends in Theory and Practice of Informatics (SOFSEM-98), Slovakia (1998).
10. Christofides, T.C., Maximal inequalities for demimartingales and a strong law of large numbers, Statistics and Probability Letters, 50, 357-363 (2000).
11. Crowder, M., Maximum likelihood estimation for dependent observations, J. Roy. Stat. Soc. B, 38, 45-53 (1975).
12. Crowder, M., On constrained maximum likelihood estimation with non-iid. observations, Ann. Inst. Statist. Math., 36, 239-249 (1983).
13. Figueira, J., Greco, S., Erhgott, M. (eds), Multiple criteria decision analysis - State of the art - Survey, 667-708. Springer's International Series, Springer, New York (2005).
14. Furrer, R., M-Estimation for dependent random variables, Stat. Prob. Letters, 57, 337-341 (2002).
15. Glynn, P.W., Rubino, G., Tuffin, B., Robustness properties and confidence interval reliability. In *Rare Event Simulation*, G. Rubino & B. Tuffin (eds). Wiley, chap. 4, (2009).
16. Hurtado, J.E., An examination of methods for approximating implicit limit state functions from the viewpoint of statistical learning theory, Structural Safety, 26, 271-293 (2004).
17. Knowles, J., Corne, D., Properties of an adaptive archiving algorithm for storing nondominated vectors, IEEE Trans. Evol. Comput., 7, 100-116 (2003).
18. Kontorovich, L.A., Ramanan, K., Concentration inequalities for dependent random variables via the martingale method, The Annals of Probability, 36, 2126-2158 (2008).
19. Kroese, D.P., Rubinstein, R.Y., Simulation and the Monte Carlo Method (2nd edition), Wiley (2007).
20. Lang, B., Monotonic multi-layer perceptron networks as universal approximators. In: W. Duch et al. (eds.): ICANN 2005, Lecture Notes in Computer Science, 3697, 31-37 (2005).
21. L'Ecuyer, P., Demers, V., Tuffin, B., Rare events, splitting, and quasi-Monte Carlo, ACM Transactions on Modeling and Computer Simulation (TOMACS), 17, Art. 9 (2007).
22. Lemaire, M., Pendola, M., PHIMECA-SOFT, Structural Safety, 28, 130-149 (2006).
23. Li, H.S., Lü, Z.Z., Yue, Z.F., Support vector machine for structural reliability analysis, Applied Mathematics and Mechanics, 27, 1295-1303 (2006).
24. Lin, D.K.J., A new class of supersaturated design, Technometrics, 35, 28-31 (1993).
25. van Leeuwen, J., Wood, D., The measure problem for rectangular ranges in $d-$space, Journal of Algorithms, 2, 282-300 (1981).
26. Limbourg, Ph., de Rocquigny, E., Andrianov, G., Accelerated uncertainty propagation in two-level probabilistic studies under monotony, Reliab. Eng. Syst. Safety, 95, 998-1010 (2010).
27. MacKay, M.D., Beckman, R.J., Conover, W.J., A comparison of three methods for selecting values of input variables in the analysis of output from a computer code, Technometrics, 21, 239-245 (1979).
28. Madsen, H.O. , Ditlevsen, O., Strutural reliability methods. Wiley (1996).
29. Munoz-Muniga, M., Garnier, J., Remy, E., de Rocquigny, E., Adaptative Directional Stratification for controlled estimation of the probability of a rare event, submitted to Reliab. Eng. Syst. Safety (2010).
30. O'Hagan, A., Bayesian analysis of computer code outputs: a tutorial, *Reliab. Engin. Sys. Safety*, 91, 1290-1300 (2006).
31. Overmars, M. H., Yap, C.-K., New upper bounds in Klee's measure problem, SIAM Journal of Computing 20(6), 1034-1045 (1991).
32. Pelckmans, K., Espinoza, M., De Brabanter, J., Suykens, J.A.K., De Moor, B., Primal-dual monotone kernel regression, Neural Processing Letters, 22, 171-182 (2005).
33. Rajabalinejad, M., Meester, L.E., van Gelder, P.H.A.J.M., Vrijling, J.K., Dynamic bounds coupled with Monte Carlo simulations, Reliab. Eng. Syst. Safety, *in press* (2010).
34. Ranjan, P., Bingham, D., Michailidis, G., Sequential experiment design for contour estimation from complex computer codes, Technometrics, 50, 527-541 (2008).
35. de Rocquigny, E., Devictor, N., Tarantola, S. (eds), Uncertainty in Industrial Practice. A guide to quantitative uncertainty management, Wiley, chap. 18, 259-281 (2008).
36. 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, 363-374 (2009).
37. Rubino, G., Tuffin, B., Introduction to rare event simulation. In *Rare Event Simulation*, G. Rubino & B. Tuffin (eds). Wiley, chap. 1, (2009).
38. Rüschendorf, L., On the distributional transform, Sklar's Theorem, and the empirical copula process, J. Statist. Plan. Infer., 139, 3921-3927 (2009).
39. Schmidt, K.D., On the covariance of monotone functions of a random variable, Report of Technische Universität Dresden, URL *www.math. tu-dresden.de/sto/schmidt/dsvm/dsvm2003-4.pdf* (2003).
40. Shan, S., Garywang, G., Failure surface frontier for reliability assessment on expensive performance function, Journal of Mechanical Design, 128, 1227-1235 (2006).
41. Wilkinson, L., The Future of Statistical Computing, Technometrics, 50(4), 418-435 (2008).

A Space transformation

When inputs $\mathbf{X} = (X_1, \ldots, X_d)$ are independent and continuous, $\Psi_{F_\mathbf{X}}^{-1}$ is simply the product of inverse cumulative distribution functions (cdf), and the item **(2)** in Assumption 3 (§ 2.2.1) is satisfied since univariate cdf preserve monotony. In dependent cases (and possibly when inputs mix continuous and discrete distributions), the generalized Rosenblatt's transform (Rüschendorf 2009) can be used if the inputs can be stochastically conditioned, e.g.:

$$F_\mathbf{X}(x_1, \ldots, x_d) = F_1(x_1) \prod_{k=2}^{d} F_{k|1,\ldots,k-1}(x_k | x_1, \ldots, x_{k-1}).$$

The the item **(2)** appears trickier to satisfy: $\Psi_{F_\mathbf{X}}^{-1}(\boldsymbol{\nu})$ must be an increasing function of all components of $\boldsymbol{\nu}$. Next lemma, however, provides a sufficient condition. Finally, a practical result which ensures Condition **(2)** has been obtained by Chen (2009) in the case when the input distribution is chosen multinormal. Indeed it deserves particular attention since, in real applications, a multinormal copula is often selected as an approximate way to tackle the difficulties of assessing correlations between input parameters.

Lemma 3 *Assume that for $k = 2, \ldots, d$, there exists a mapping g_k and a set of (possibly random) parameters θ_k independent of $X_1, \ldots, X_k$ such that:*

(i) $X_k = g_k(X_1, \ldots, X_{k-1}, \theta_k)$,
(ii) *g_k is a globally increasing function of $X_1, \ldots, X_{k-1}$;*

then the inverse Rosenblatt's transform

$$\Psi_{F_{\mathbf{X}}}^{-1}(v_1,\ldots,v_d) = F_1^{-1}(v_1)\prod_{k=2}^{d} F_{k|1,\ldots,k-1}^{-1}(v_k|v_1,\ldots,v_{k-1}),$$

is a globally increasing function of $\boldsymbol{\nu}$.

Example 5 In corollary of Lemma 3, any standard binormal input distribution with positive correlation coefficient μ ensures that $\Psi_{F_{\mathbf{X}}}^{-1}(\boldsymbol{\nu})$ is increasing. Indeed, $\mathbf{X} = (X_1, X_2)$ where $X_1 \sim \mathscr{N}(0,1)$ and $X_2 = \mu X_1 + \sqrt{1-\mu^2}\theta$ with $\theta \sim \mathscr{N}(0,1)$.

Proof ***(Lemma 3)***

Assume (i). $\forall t \in \mathbb{R},\ \forall k \in \{2,\ldots,d\}$, denote $p_{\theta_k}^t(X_1,\ldots,X_{k-1}) = P(g_k(X_1,\ldots,X_{k-1},\theta_k) < t | X_1,\ldots,X_{k-1})$. Then, $\forall y \in \mathbb{R}$, let $A_{X_1,\ldots,X_{k-1}}^t(y)$ denote the event $\{p_{\theta_k}^t(X_1,\ldots,X_{k-1}) \le y\}$ defined in a probability space $(\{0,1\}, \mathscr{A}_k^t, \mathscr{P})$. Thus, by definition,

$$F_{k|1,\ldots,k-1}^{-1}(y|X_1,\ldots,X_{k-1}) = \inf\left\{t \in \mathbb{R} \,|P\left(A_{X_1,\ldots,X_{k-1}}^t(y)\right) = 1\right\}.$$

Assuming (ii), $p_{\theta_k}^t(X_1,\ldots,X_{k-1})$ is a globally decreasing function of $X_1,\ldots,X_{k-1}$. Thus the occurence of event $A_{X_1,\ldots,X_{k-1}}^t(y)$ similarly decreases. It also appears necessary to increase t in order to counterbalance this tendency, which implies that the minimum value of all $t \in \mathbb{R}$ such that $\mathscr{P}(A_{X_1,\ldots,X_{k-1}}^t(y)) = 1$ increases. Hence $F_{k|1,\ldots,k-1}^{-1}$ is a globally increasing function of $X_1,\ldots,X_{k-1}$, $\forall k \in \{2,\ldots,d\}$. Since $X_1 = F^{-1}(v_1)$ is naturally an increasing function of v_1, a simple recursive reasoning shows that $F_{k|1,\ldots,k-1}^{-1}$ is a globally increasing function of $v_1,\ldots,v_{k-1}$. The statement of the lemma follows.

B Computation of progressive bounds

A key point is that any bound computation does not need any call to the costly function G. One must evaluate the d-dimensional integral of a known function (the joint pdf of $\mathbf{X}$) over the dominated spaces. The space transformation proposed in § 2.2.1 allows for a simplifed and exact computation of the bounds.

Consider first the simple case described in Lemma 1 where dominated spaces $(\mathbb{D}_{\tilde{\mathbf{x}}}^-, \mathbb{D}_{\tilde{\mathbf{x}}}^+)$ are defined by a single point $\tilde{\mathbf{x}} \in \mathbb{D}$. Denote $\tilde{\boldsymbol{\nu}} = (\tilde{\nu}_1,\ldots,\tilde{\nu}_d)$ its image point into $\mathbb{U}$ through $\Psi_{F_{\mathbf{X}}}^{-1}$. Denote similarly $(\mathbb{U}_{\tilde{\boldsymbol{\nu}}}^-, \mathbb{U}_{\tilde{\boldsymbol{\nu}}}^+)$ the image spaces of $(\mathbb{D}_{\tilde{\mathbf{x}}}^-, \mathbb{D}_{\tilde{\mathbf{x}}}^-)$ through $\Psi_{F_{\mathbf{X}}}^{-1}$.

If $\tilde{G}(\tilde{\boldsymbol{\nu}}) \le 0$, the failure-dominated image space $\mathbb{U}_{\tilde{\boldsymbol{\nu}}}^-$ is a hypercube whose vertexes are the origin $0^d = (0,\ldots,0)$, the point $\tilde{\boldsymbol{\nu}}$ and the projections of $\tilde{\boldsymbol{\nu}}$ on axes of the canonical base $\mathscr{B}_c(\mathbb{U})$. Then $\mathbb{U}_{\tilde{\boldsymbol{\nu}}}^+ = \{1^d\} = \{(1,\ldots,1)\}$ and

$$P\left(\mathbf{X} \in \mathbb{D}_{\tilde{\mathbf{x}}}^+\right) = \int_{\mathbb{D}} \prod_{i=1}^{d} \mathbb{1}_{\{x_i \le \tilde{x}_i\}} f_{\mathbf{X}}(\mathbf{x})\, d\mathbf{x},$$

$$= P\left(\boldsymbol{\nu} \in \mathbb{U}_{\tilde{\boldsymbol{\nu}}}^+\right) = \prod_{i=1}^{d}(1-\tilde{\nu}_i).$$

If now $\tilde{G}(\tilde{\boldsymbol{\nu}}) > 0$, the safety-dominated image space $\mathbb{U}_{\tilde{\boldsymbol{\nu}}}^+$ is a hypercube whose vertexes are the opposite corner 1^d, the point $\tilde{\boldsymbol{\nu}}$ and the axes of the base resulting from the $\pi-$rotation of $\mathscr{B}_c(\mathbb{U})$, centered on 1^d. Then $\mathbb{U}_{\tilde{\boldsymbol{\nu}}}^- = \{0^d\}$ and

$$P\left(\mathbf{X} \in \mathbb{D}_{\tilde{\mathbf{x}}}^-\right) = P\left(\boldsymbol{\nu} \in \mathbb{U}_{\tilde{\boldsymbol{\nu}}}^+\right) = \prod_{i=1}^{d} \tilde{\nu}_i.$$

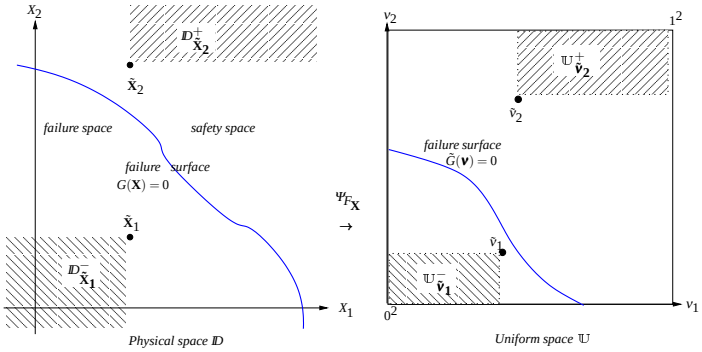

Fig. 10 A two-dimensional case. Axes define canonical bases. Transforming physical space $\mathbb{D}$ in uniform space $\mathbb{U}$ simplifies the calculus of $P(\mathbf{X} \in \mathbb{D}_{\tilde{\mathbf{x}}_1}^-)$ and $P(\mathbf{X} \in \mathbb{D}_{\tilde{\mathbf{x}}_2}^+)$.

The calculus of probability bounds is thus immediate in this simple case. An example involving two points in failure and safety spaces, respectively, is displayed on Figure 10.

The more complicated case of computing bounds (1), emanating from a progressive DoE (for instance as displayed on Figure 1), can be treated in two ways.

First and fortunately, an exact method can be implemented. Calculating the volume of a union of hyperrectangles sharing the same orthogonal basis is known in computational geometry as Klee's measure problem, for which recursive sweepline algorithms (van Leeuwen and Wood 1981) can provide exact solutions. Overmars and Yap (1991) and Chlebus (1998) proposed algorithms with complexity $O(n^{d/2}\log n)$ for $d \ge 3$. In comparison, typical algorithms used in MOO (cf. Remark 1) have complexity $O(n^{d+1})$ (Knowles and Corne 2003). Although some recent improvements have been brought by Chan (2008), Overmars and Yap's complexity has not significantly decreased since[2].

When d exceeds 4 or 5, this exact method rapidly appears too costly. Therefore a trivial MC method can take over. Volume V_n^- (*resp.* V_n^+) can be consistently estimated by sampling uniformly M independent items $\boldsymbol{\nu}_1,\ldots,\boldsymbol{\nu}_M \in \mathbb{U}^M$ then computing the proportion of items inside $\mathbb{U}_n^-$ (*resp.* $\mathbb{U}/\mathbb{U}_n^+$). The sampling cost is here independent of d. However, this statistical approach introduces estimation error which weakens one of the main property of MRM algorithms: *computed* bounds are not certain 100%-confidence bounds for p_f. Since in practice both methods suffers from numerical roundness in $[0,1]^d$, M should be chosen high enough to merge with the roundness error encountered using the sweepline in small dimensions.

C Proofs of statements in the main text

In next statements and proofs, the notation $\mathscr{F}_n$ is often removed for a better reading. Especially, the expectancies and variances of quantities θ_n are defined with respect to the available $\sigma-$algebra $\mathscr{F}_n$.

Proof ***(Proposition 1)***

[2] Additional details and the pseudo-code of a sweepline algorithm with exponential complexity, used in the experiments presented in this article, are available in the Electronic Supplementary Material.

One may write $\ell_n''(p_f) = \sum_{k=1}^n \tilde{\omega}_k(p_f) S_k(p_f)$ with

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

Note that $\forall k$, $p_k \in \{p_{k-1}^-, p_{k-1}^+\}$. If $p_k = p_{k-1}^+$, then $S_k(p_f) = -(p_{k-1}^+ - p_f)/(p_f - p_{k-1}^-) \leq 0$ if $p_f \in (p_{k-1}^-, p_{k-1}^+) \supset [p_{n-1}^-, p_{n-1}^+]$. Similarly, if $p_k = p_{k-1}^-$, then $S_k(p_f) = -(p_f - p_{k-1}^-)/(p_{k-1}^+ - p_f) < 0$ if $p_f \in (p_{k-1}^-, p_{k-1}^+)$. Hence $\ell_n''(p_f) < 0$ in (p_{n-1}^-, p_{n-1}^+). Besides, note that $\lim_{p_f \to p_{n-1}^-} \ell_n'(p_f) = \infty$ and $\lim_{p_f \to p_{n-1}^+} \ell_n'(p_f) = -\infty$. Hence, by twice continuity and differentiability of $\ell_n(p_f)$, the mean value theorem implies the existence and unicity of a MLE $\hat{p_f}_n$ in (p_{n-1}^-, p_{n-1}^+).

Proof (**Theorem 1**)

We first prove the following result: $\forall\, n \geq 1$,

$$J_n^{-1}(p_f) \leq V_n^{MC}(p_f)\left(\frac{2p_f}{1-p_f}\right)\left(\rho_{n-1}^{med}\right)^2. \tag{19}$$

Indeed, since $\tilde{\omega}_{n+1}$ depends only on $\mathscr{F}_n$ for $n \geq 1$, notice that

$$\mathrm{E}_{\mathscr{F}_n}\left[\ell_n'(p_f)\right] = \sum_{k=1}^n \mathrm{E}\left[\tilde{\omega}_k(p_f)\mathrm{E}\left[p_k - p_f|\mathscr{F}_{k-1}\right]\right] = 0.$$

so that the Fisher information $J_n(p_f) = \mathrm{Var}[\ell_n'^2(p_f)] = \mathrm{E}[\ell_n'^2(p_f)]$ is equal to $-\mathrm{E}[\ell_n''(p_f)]$ by twice differentiability and continuity of $\ell_n(\cdot)$, like in a classic iid. case. Condition (i) implies that $\forall n < \infty$, $p_{n-1}^- < p_f < p_{n-1}^+$, ie. p_f cannot be reached in any finite number of iterations, so that these quantities are well defined. Using the notation $S_k(p_f)$ defined in the proof of Proposition 1,

$$\begin{aligned} J_n(p_f) &= -\mathrm{E}_{\mathscr{F}_n}\left[\frac{p_f(1-p_f)}{nV_n^{MC}(p_f)}\sum_{k=1}^n \tilde{\omega}_k(p_f) S_k(p_f)\right], \\ &= n^{-1}\frac{\tilde{J}_n(p_f)}{V_n^{MC}(p_f)} \end{aligned}$$

with $\tilde{J}_n(p_f)$ equal to

$$\mathrm{E}_{\mathscr{F}_n}\left[\sum_{k=1}^n p_f(1-p_f)\left(p_{k-1}^+ - p_f\right)^{2\xi_{\mathbf{v}_k}-2}\left(p_f - p_{k-1}^-\right)^{-2\xi_{\mathbf{v}_k}}\right].$$

Thus

$$\begin{aligned} \tilde{J}_n(p_f) = \sum_{k=1}^n \Bigg\{ &\mathrm{E}_{\mathscr{F}_k}\left[\xi_{\mathbf{v}_k}\left(\frac{p_f(1-p_f)}{\left(p_f - p_{k-1}^-\right)^2}\right)\right] \\ &+ \mathrm{E}_{\mathscr{F}_n}\left[(1-\xi_{\mathbf{v}_k})\left(\frac{p_f(1-p_f)}{\left(p_{k-1}^+ - p_f\right)^2}\right)\right]\Bigg\}. \end{aligned}$$

Since

$$\frac{1}{p_f}\left(\left(p_{k-1}^+ - p_f\right) + \left(p_f - p_{k-1}^-\right)\right) \leq \rho_{k-1},$$

hence

$$\begin{aligned} p_{k-1}^+ - p_f &\leq p_f(\rho_{k-1} - 1) + p_{k-1}^- \leq p_f(\rho_{k-1} - 1) + p_f, \\ p_f - p_{k-1}^- &\leq p_f(\rho_{k-1} + 1) - p_{k-1}^+ \leq p_f(\rho_{k-1} + 1). \end{aligned}$$

Thus

$$\begin{aligned} \frac{p_f(1-p_f)}{\left(p_{k-1}^+ - p_f\right)^2} &\geq \frac{1-p_f}{p_f\rho_{k-1}^2}, \\ \frac{p_f(1-p_f)}{\left(p_f - p_{k-1}^-\right)^2} &\geq \frac{1-p_f}{p_f(1+\rho_{k-1}^2)}. \end{aligned}$$

Consequently, since $(\rho_n)_n$ is a strictly decreasing process,

$$\tilde{J}_n(p_f) \geq \left(\frac{1-p_f}{p_f}\right)\mathrm{E}_{\mathscr{F}_n}\left[\frac{1}{\rho_{n-1}^2}\left(T_{1,n} + T_{2,n}\left(1 + 1/\rho_{n-1}\right)^{-2}\right)\right]$$

where $T_{1,n} = \sum_{k=1}^n (1 - \xi_{\mathbf{v}_k})$ and $T_{2,n} = \sum_{k=1}^n \xi_{\mathbf{v}_k}$. Thus

$$\begin{aligned} \tilde{J}_n(p_f) &\geq n\left(\frac{1-p_f}{p_f}\right)\mathrm{E}_{\mathscr{F}_n}\left[\rho_{n-1}^{-2}\right], \\ &\geq n\left(\frac{1-p_f}{p_f}\right)\left(\rho_{n-1}^{-2}\right)^{med} P\left(\rho_{n-1}^{-2} \leq \left(\rho_{n-1}^{-2}\right)^{med}\right) \end{aligned}$$

using Markov's inequality, hence $\tilde{J}_n(p_f) \geq n\left(\frac{1-p_f}{2p_f}\right)\left(\rho_{n-1}^{med}\right)^{-2}$, which proves (19).

The weak consistency and the asymptotic normality of the MLE can be proved using arguments mainly developed by Crowder (1975,1983). We first obtain the conditions of a martingale central limit theorem, then the weak consistency. Both results can then be combined to lead to the final asymptotic normality.

First consider the fact that $(\ell_n'(p_f))_n$ is a martingale:

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

Furthermore, because of the increasing of the weights,

$$J_n(p_f) = \sum_{k=1}^n \mathrm{E}\left[\tilde{\omega}_k(p_f)\right] \leq n\mathrm{E}\left[\tilde{\omega}_n(p_f)\right] \leq n \sup_{\{\mathscr{F}_n^u\}} \tilde{\omega}_n(p_f)$$

where $\{\mathscr{F}_n^u\}$ describes the family of all possible $\sigma-$algebrae generated by the all possible n successions of nested uniform samplings. Condition (i) implies this upper bound is finite. Therefore $(\ell_n'(p_f))_n$ is a square integrable martingale. One can show that

$$\Delta_n^2(p_f) = \left(\ell_n'(p_f) - \ell_{n-1}'(p_f)\right)^2 = \tilde{\omega}_n^2(p_f)(p_n - p_f)^2$$

hence

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

since

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

Thus, the increasing (or *angle bracket*) process linked to $(\ell_n'(p_f))_n$, defined by

$$< \ell'(p) >_n = \sum_{k=1}^n \mathrm{E}\left[\Delta_k^2(p_f)|\mathscr{F}_{k-1}\right],$$

is such that $J_n(p_f) = \mathrm{E}\left[< \ell'(p) >_n\right]$, and from (19) $\exists\, C > 0$ such that

$$\left|\frac{< \ell'(p) >_n}{J_n(p_f)} - 1\right| \leq \frac{C\left(\rho_{n-1}^{med}\right)^2}{n}\left|\sum_{k=1}^n \left(\tilde{\omega}_k(p_f) - \mathrm{E}_{\mathscr{F}_k}\left[\tilde{\omega}_k(p_f)\right]\right)\right| \tag{20}$$

which tends to 0 under Condition (ii), then

$$\frac{< \ell'(p) >_n}{J_n(p_f)} \xrightarrow[n\geq 1]{a.s.} 1. \tag{21}$$

Besides, for any $\varepsilon > 0$,

$$P\left(\frac{|\Delta_k(p_f)|}{\sqrt{J_n(p_f)}} < \varepsilon\right) = P\left(A^{\varepsilon}_{n,k}(p_f) < \xi_{\nu_k} < B^{\varepsilon}_{n,k}(p_f)\right)$$

with ξ_{ν_k} a Bernoulli variable in $\{0,1\}$ and

$$A^{\varepsilon}_{n,k}(p_f) = \frac{p_f - p^-_{k-1} - \varepsilon\sqrt{J_n(p_f)}}{p^+_{k-1} - p^-_{k-1}},$$
$$B^{\varepsilon}_{n,k}(p_f) = \frac{p_f - p^-_{k-1} + \varepsilon\sqrt{J_n(p_f)}}{p^+_{k-1} - p^-_{k-1}}.$$

which are defined for all k under Condition (i). With $J_n(p_f) \to \infty$ and $p^+_{n-1} - p^-_{n-1}$ tending to a finite positive or null limit, then

$$\lim_{n\to\infty} P\left(\frac{|\Delta_k(p_f)|}{\sqrt{J_n(p_f)}} < \varepsilon\right) = P\left(-\infty < \xi_{\nu_k} < \infty\right) = 1$$

so $\frac{|\Delta_k(p_f)|}{\sqrt{J_n(p_f)}} \xrightarrow[n\to\infty]{I\!P} 0$. Thus, for any $\varepsilon > 0$, $\exists\ N_\varepsilon \in I\!N^*$, $N^* < \infty$, such that $\forall\ n > N^*$ no event $|\Delta_k(p_f)| \geq \varepsilon\sqrt{J_n(p_f)}$ can be observed. Necessarily, the sum $\sum_{k=1}^n J_k^{-1}(p_f)\Delta_k^2(p_f)\mathbb{1}_{\left\{|\Delta_k(p_f)| \geq \varepsilon\sqrt{J_n(p_f)}\right\}}$ is bounded. Since $J_n(p_f) \to \infty$, a Lindeberg condition can be satisfied:

$$\frac{1}{J_n(p_f)}\sum_{k=1}^n \mathrm{E}\left[\Delta_k^2(p_f)\mathbb{1}_{\left\{\frac{|\Delta_k(p_f)|}{\sqrt{J_n(p_f)}} \geq \varepsilon\right\}} \Big| \mathscr{F}_{k-1}\right] \xrightarrow[n\to\infty]{I\!P} 0. \tag{22}$$

Finally (21) and (22) prove the martingale central limit theorems (Bercu 2008):

$$J_n^{-1/2}(p)\ell'_n(p_f) \xrightarrow[n\to\infty]{\mathscr{L}} \mathscr{N}(0,1), \tag{23}$$

$$\frac{\sqrt{J_n(p_f)}}{< \ell'(p_f) >_n}\ell'_n(p_f) \xrightarrow[n\to\infty]{\mathscr{L}} \mathscr{N}(0,1). \tag{24}$$

The weak consistency of the MLE can be proved using a sufficient condition given by Crowder (1975). Denote

$$\Gamma_n = \inf_{\{\mathscr{F}_n\}}\left\{p_f - p^-_n, p^+_n - p_f\right\}$$

and $\sigma_n = \sqrt{n}/(\Gamma^2_{n+1}\rho^{\mathrm{med}}_{n-1})$. Clearly the sequence $\{\sigma_n\}$ is positive and tends to infinity, whatever p_f. Denoting

$$M_n = \sigma_n^{-1}\left(p^+_{n-1} - p^-_{n-1}\right)^{-2}\sqrt{\sum_{k=1}^n\left(\tilde{\omega}_k(p_f) - \mathrm{E}_{\mathscr{F}_k}\left[\tilde{\omega}_k(p_f)\right]\right)}$$

one has $M_n = o(1)$. Noticing that $\forall\ n \geq 0$, $-S_n(p_f) = \tilde{\omega}_n(p_f)(p_f - p_n)^2 \geq \tilde{\omega}_n(p_f)\Gamma^2_{n+1}$, one has

$$-\sigma_n\frac{\ell''_n(p_f)}{\sqrt{J_n(p_f)}} \geq \frac{\sum_{k=1}^n \tilde{\omega}_k^2(p_f)}{\sqrt{M_n^2 + n^{-1}(\rho^{\mathrm{med}}_{n-1})^2\sum_{k=1}^n \tilde{\omega}_k^2(p_f)}} \overset{\infty}{\geq} \frac{\sqrt{n}}{\rho^{\mathrm{med}}_{n-1}}$$

thus

$$P\left(-\sigma_n\frac{\ell''_n(p_f)}{\sqrt{J_n(p_f)}} \geq 1\right) \geq P\left(\sqrt{n}(\rho^{\mathrm{med}}_{n-1})^{-1}(1+o(1)) \geq 1\right)$$

which implies that, $\forall \tilde{p_f}$ in an neighborhood of the true value p_f where ℓ''_n is continuous, such that $\alpha = |\tilde{p_f} - p_f| > 0$,

$$\lim_{n\to\infty} P\left(-\sigma_n(\tilde{p_f} - p_f)\frac{\ell''_n(p_f)}{\sqrt{J_n(p_f)}}(\tilde{p_f} - p_f) \geq \alpha^2\right) = 1.$$

From Equation (2.3) in Crowder (1975), this proves the weak consistency of the MLE.

Finally, since $p^-_{n-1} < \hat{p_f}_n < p^+_{n-1}$, for any n there always exists an open neighborhood $\mathscr{V}_{\hat{p_f}_n}$ of p_f containing $\hat{p_f}_n$. From twice differentiability of $\ell_n(\cdot)$ and continuity of $\ell'_n(\cdot)$ in $\mathscr{V}_{\hat{p_f}_n}$, the mean value theorem implies there exists some intermediate point $\bar{p_n} \in \mathscr{V}_{\hat{p_f}_n}$ between p_f and $\hat{p_f}_n$ such that

$$\ell'_n\left(\hat{p_f}_n\right) = 0 = \ell'_n(p_f) + \left(\hat{p_f}_n - p_f\right)\ell''_n(\bar{p_n})$$

Thus, with $\ell''_n(\tilde{p}_n) \neq 0$,

$$\left(\hat{p_f}_n - p_f\right) = \ell'_n(p_f)\left(-\ell''_n(\bar{p_n})\right)^{-1}. \tag{25}$$

Then

$$\left|-\ell''(\bar{p_n}) - J_n(p_f)\right| \leq |A_n(p_f)| + B_n(p_f) + C_n(p_f, \bar{p_n})$$

with

$$A_n(p_f) = \sum_{k=1}^n\left(\tilde{\omega}_k(p_f) - \mathrm{E}_{\mathscr{F}_k}\left[\tilde{\omega}_k(p_f)\right]\right),$$
$$B_n(p_f) = \sum_{k=1}^n \tilde{\omega}_k(p_f)\left|\tilde{\omega}_k(p_f)(p_f - p_k)^2 - 1\right|,$$
$$C_n(p_f, \bar{p_n}) = \sum_{k=1}^n\left|V_k(p_f, \bar{p_n})\right|$$

with $V_k(p_f, \bar{p_n}) = \tilde{\omega}_k^2(\bar{p_n})(\bar{p_n} - p_k)^2 - \tilde{\omega}_k^2(p_f)(p - p_k)^2$. Since $\bar{p_f}_n \in [\hat{p_f}_n, p_f]$ then $\bar{p_n} \xrightarrow{I\!P} p_f$ by weak consistency of $\hat{p_f}_n$ and $V_k(p_f, \bar{p_n}) \xrightarrow{I\!P} 0$ by continuity of mappings $x \mapsto \tilde{\omega}_k(x)$ on $[p^-_{n-1}, p^+_{n-1}]$, using Slutsky's theorem. Then the sequence $\{|V_k(p_f, \bar{p_n})|\}$ is bounded and $\exists\ K > 0$, $K < \infty$, such that $C_n(p_f, \bar{p_n}) \leq nK$. We deduce from (19) that $\lim_{n\to\infty} J_n^{-1}(p_f)C_n(p_f) = 0$. Condition (i) implies that $\lim_{n\to\infty} J_n^{-1}(p_f)|A_n(p_f)| = 0$. Then it is worthy to note that because of Condition (iii),

$$\tilde{\omega}_k(p_f)(p_f - p_k)^2 - 1 \xrightarrow[n\to\infty]{I\!P} 0.$$

Denoting $\mu_n = \sup_{k\leq n}|\tilde{\omega}_k(p_f)(p_f - p_k)^2 - 1|$, the sequence $\{\mu_n\}$ decreases toward 0 and

$$\frac{B_n(p_f)}{J_n(p_f)} = \left(\frac{B_n(p_f)}{< \ell'(p_f) >_n}\right)\left(\frac{< \ell'(p_f) >_n}{J_n(p_f)}\right),$$
$$\leq \mu_n\left(\frac{< \ell'(p_f) >_n}{J_n(p_f)}\right)$$

which tends to 0 due to (21). Finally,

$$\frac{-\ell''(\bar{p_n})}{J_n(p_f)} \xrightarrow[n\to\infty]{I\!P} 1. \tag{26}$$

Combining the results (25), (23) and (26) and taking account of the consistency of $\hat{p_f}_n$, the final result can finally be concluded from Theorem 3 in Crowder (1983).

Proof ***(Proposition 2)***

Remind that $V_n^{MLE} = J_n^{-1}(p_f)$ and note that, taking back to the notations of the previous proof, $\hat{J}_n(p_f) = < \ell'(p_f) >$. By twice continuity and derivability of $\hat{J}^{-1}(.)$ in (p^-_n, p^+_n), a Taylor expansion gives

$$\hat{J}^{-1}(\hat{p_f}_n) = \hat{J}^{-1}(p_f) - \frac{\hat{J}'_n(p_f)}{\hat{J}_n^2(p_f)}\left(\hat{p_f}_n - p_f\right)(1+o(1)).$$

Thus, after some calculus,

$$\frac{\hat{J}_n^{5/2}(p_f)}{|\hat{J}_n'(p_f)|}\left(\hat{f}^{-1}(\hat{p_f}_n)-V_n^{MLE}\right) = W_nY_n+W_nZ_n$$

with

$$W_n = \sqrt{\frac{J_n(p_f)}{\hat{J}_n(p_f)}},$$

$$U_n = \mathrm{sgn}(\hat{J}_n'(p_f))\sqrt{J_n(p_f)}\left(\hat{p_f}_n-p_f\right)(1+o(1)),$$

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

From (21) $W_n \xrightarrow{a.s.} 1$, and $U_n \xrightarrow{\mathscr{L}} \mathscr{N}(0,1)$ from Theorem 1. 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

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

which is always nonzero under Assumption (v). Using Hölder's inequality, one has

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

hence

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

Using Hölder's inequality again, one has

$$\frac{\hat{J}_n^{3/2}(p_f)}{|\hat{J}_n'(p_f)|} \le \sqrt{\sum_{k=1}^{n}\frac{\left(p_{k-1}^+-p_f\right)\left(p_f-p_{k-1}^-\right)}{2p_f-\left(p_{k-1}^+ + p_{k-1}^-\right)}}$$

and it can be shown with simple calculus that each term of the sum is stricly smaller than 1. Then

$$|Z_n| \le \sqrt{n}\left|\frac{\hat{J}_n(p_f)}{J_n(p_f)}-1\right|.$$

and from (20) and (19), $\exists D>0$, $D<\infty$, such that

$$|Z_n| \le Dn^{1/2-\delta}$$

which converges toward 0 if $\delta > 1/2$. Thus $Z_n \xrightarrow{a.s.,I\!P} 0$.

Proof ***(Lemma 2)***

$$\begin{aligned}\mathrm{E}_{\mathscr{F}_k}[\tilde{p}_k] &= \mathrm{E}\left[p_{k-1}^- + \mathrm{E}\left[\frac{\xi_{\boldsymbol{\nu}_k}}{f_{k-1}(\boldsymbol{\nu}_k)}\middle|\mathscr{F}_{k-1}\right]\right],\\ &= \mathrm{E}\left[p_{k-1}^- + \int_{\mathbb{U}_{k-1}} P\left(\tilde{G}(\boldsymbol{\nu})\le 0\right)\frac{f_{k-1}(\boldsymbol{\nu})}{f_{k-1}(\boldsymbol{\nu})}\,d\boldsymbol{\nu}\right],\end{aligned}$$

and from (3), $P(\tilde{G}(\boldsymbol{\nu})\le 0|\boldsymbol{\nu}\in\mathbb{U}_{k-1}) = (p_f-p_{k-1}^-)/(p_{k-1}^+-p_{k-1}^-)$. Then

$$\begin{aligned}\mathrm{E}_{\mathscr{F}_k}[\tilde{p}_k] &= \mathrm{E}_{\mathscr{F}_{k-1}}\left[p_{k-1}^- + \frac{p_f-p_{k-1}^-}{p_{k-1}^+-p_{k-1}^-}\mathrm{Vol}(\mathbb{U}_{k-1})\right],\\ &= \mathrm{E}\left[p_{k-1}^- + \frac{p_f-p_{k-1}^-}{p_{k-1}^+-p_{i-1}^-}\left(p_{k-1}^+-p_{k-1}^-\right)\right],\\ &= p_f.\end{aligned}$$

Proof ***(Proposition 3)***

Denote $\check{V}_k = \mathrm{Var}\left[\mathrm{E}\left[\check{p}_{f\,k+1}|\mathscr{F}_k\right]\right]$. Then

$$\begin{aligned}\check{V}_{n-1} &= V_{\mathscr{F}_{n-1}}\left[\sum_{i=1}^{n-1}\omega_i\tilde{p}_i\right],\\ &= \mathrm{E}\left[\mathrm{Var}\left[\sum_{i=1}^{n-1}\omega_i\tilde{p}_i|\mathscr{F}_{n-2}\right]\right]+\check{V}_{n-2},\\ &= \sum_{k=1}^{n-1}\mathrm{E}\left[\mathrm{Var}\left[\sum_{j=1}^{k}\omega_j\tilde{p}_j|\mathscr{F}_{k-1}\right]\right].\end{aligned}$$

Since

$$\mathrm{Var}\left[\check{p}_{f\,n}\right] = \check{V}_{n-1}+\mathrm{E}\left[\mathrm{Var}\left[\sum_{j=1}^{n}\omega_j\tilde{p}_j|\mathscr{F}_{n-1}\right]\right],$$

$$\begin{aligned}\text{then}\quad \mathrm{Var}\left[\check{p}_{f\,n}\right] &= \sum_{k=1}^{n}\mathrm{E}\left[\mathrm{Var}\left[\sum_{j=1}^{k}\omega_j\tilde{p}_j|\mathscr{F}_{k-1}\right]\right],\\ &= \sum_{k=1}^{n}\mathrm{E}\left[\omega_k^2\mathrm{Var}[\tilde{p}_k|\mathscr{F}_{k-1}]\right].\end{aligned} \tag{27}$$

With $\xi_{\boldsymbol{\mu}_k}^2 = \xi_{\boldsymbol{\mu}_k}$ and since $\tilde{p}_k$ is unbiased knowing $\mathscr{F}_{k-1}$, one has

$$\begin{aligned}\mathrm{Var}[\tilde{p}_k|\mathscr{F}_{k-1}] &= E\left[\tilde{p}_k^2|\mathscr{F}_{k-1}\right]-p_f^2,\\ &= (p_{k-1}^-)^2+2p_{k-1}^-(p_f-p_{k-1}^-)\\ &\quad+\int_{\mathbb{U}_{k-1}}\frac{\xi_{\boldsymbol{\nu}}}{f_{k-1}^2(\boldsymbol{\nu})}f_{k-1}(\boldsymbol{\nu})\,d\boldsymbol{\nu}-p_f^2,\\ &= -\left(p_f^2-2p_fp_{k-1}^-+(p_{k-1}^-)^2\right)\\ &\quad+\left(p_{k-1}^+-p_{k-1}^-\right)\int_{\mathbb{U}_{k-1}}\xi_{\boldsymbol{\nu}}\frac{g_{k-1}(\boldsymbol{\nu})}{f_{k-1}(\boldsymbol{\nu})}\,d\boldsymbol{\nu}\end{aligned}$$

which is equal to

$$\left(p_{k-1}^+-p_f\right)\left(p_f-p_{k-1}^-\right)+\left(p_{k-1}^+-p_{k-1}^-\right)A_{k-1}(f_{k-1}) \tag{28}$$

where, $\forall k\ge 1$,

$$A_{k-1}(f_{k-1}) = \int_{\mathbb{U}_{k-1}}\mathbb{1}_{\{\tilde{G}(\boldsymbol{\nu})\le 0\}}\frac{g_{k-1}(\boldsymbol{\nu})}{f_{k-1}(\boldsymbol{\nu})}\,d\boldsymbol{\nu}-(p_f-p_{nk-1}^-)$$

and g_{k-1} is the uniform pdf on $\mathbb{U}_{k-1}$. Then simple algebra leads to the developed formula for V_n^{WISE} from (27) and (28), noticing that

$$\mathrm{E}\left[\left(p_{k-1}^+-p_f\right)\left(p_f-p_{k-1}^-\right)\right] = p_f(1-p_f)(1-c_{k-1}).$$

Corollary 1 is straightforward using a Lagrangian method.

Proof (**Proposition 4**)

For any $k \geq 1$ such that $b_{k-1} \leq 0$, $V_n^{*WISE} \leq V_n^{\mathrm{U}}(p_f)$. Since $(\mathbb{U}_0^+, \mathbb{U}_0^-) = (\mathbb{U}_N^+, \mathbb{U}_N^-)$ with $N \geq k_0$ then (p_0^+, p_0^-) are not trivial ($p_0^+ < 1$ and $p_0^- > 0$). Consequently, $\rho_0 < \infty$. Note that (ρ_n) is a strictly decreasing stochastic process, upperly bounded by the deterministic finite upper bound ρ_0. Note that

$$\forall k = 0, \ldots, n-1, \quad 1 - c_k \;\leq\; \frac{p_f}{1-p_f} \mathrm{E}_{\mathscr{F}_k}[\rho_k].$$

This can be proved as follows. One has $\Upsilon_k(p_f) \leq \rho_k$, thus $1 - p_k^+ \geq 1 - p_k^+ - \rho_k p_f$. Hence

$$p_f(1-p_k^+) + (1-p_f)p_k^- \;\geq\; (p_f + p_k^-) - 2p_f p_k^- - \rho_k p_f^2$$

then

$$\begin{aligned} p_f(1-p_k^+) + (1-p_f)p_k^- - (1-p_k^+)p_k^-, & \\ \geq (p_f + p_k^-) - 2p_f p_k^- - \rho_k p_f^2 - p_k^- + p_k^+ p_k^-, & \\ \geq p_f - 2p_f p_k^- - \rho_k p_f^2 + p_f p_k^- \quad \text{since } p_f \leq p_k^+, & \\ \geq p_f - \rho_k p_f^2 - p_f^2 \quad \text{since } p_f \geq p_k^-, & \\ \geq p_f(1-p_f) - \rho_k p_f^2. & \end{aligned}$$

Thus

$$\frac{p_k^-}{p_f} + \frac{1-p_k^+}{1-p_f} - \frac{p_k^-(1-p_k^+)}{p_f(1-p_f)} \leq 1 - \frac{p_f \rho_k}{1-p_f}$$

and the result follows by taking the expectation over $\boldsymbol{\nu}_1, \ldots, \boldsymbol{\nu}_k$.

Finally, we obtain

$$V_n^{\mathrm{U}}(p_f) \leq \frac{p_f(1-p_f)}{n} \frac{1}{n} \sum_{i=1}^{n} \omega_i^2 \frac{p_f \rho_0}{1-p_f}$$

which proves (11) after basic algebra.

Proof (**Proposition 5**)

Denoting $f_{n-1}(\boldsymbol{\nu}) = k(\boldsymbol{\nu}) / \int_{\mathbb{U}_{n-1}} k(\boldsymbol{\nu})\, d\boldsymbol{\nu}$ where k is a bounded function from $\mathbb{U}_{n-1}$ to $\mathbb{R}_+^*$, the Cauchy-Schwarz inequality proves that

$$\begin{aligned} A_{n-1}(f_{n-1}) \;=\; & \left[\int_{\mathbb{U}_{n-1}} \frac{\xi_{\boldsymbol{\nu}} g_{n-1}(\boldsymbol{\nu})}{k(\boldsymbol{\nu})}\, d\boldsymbol{\nu} \right] \left[\int_{\mathbb{U}_{n-1}} k(\boldsymbol{\nu}) g_{n-1}(\boldsymbol{\nu})\, d\boldsymbol{\nu} \right] \\ & \times \left(p_{n-1}^+ - p_{n-1}^-\right) - \left(p_f - p_{n-1}^-\right), \\ \geq \; & \left[\int_{\mathbb{U}_{n-1}} \xi_{\boldsymbol{\nu}} g_{n-1}(\boldsymbol{\nu})\, d\boldsymbol{\nu} \right]^2 \left(p_{n-1}^+ - p_{n-1}^-\right) \\ & - \left(p_f - p_{n-1}^-\right), \\ \geq \; & \left(\frac{p_f - p_{n-1}^-}{p_{n-1}^+ - p_{n-1}^-} \right)^2 \left(p_{n-1}^+ - p_{n-1}^-\right) - \left(p_f - p_{n-1}^-\right), \\ \geq \; & - \frac{\left(p_{n-1}^+ - p_f\right)\left(p_f - p_{n-1}^-\right)}{\left(p_{n-1}^+ - p_{n-1}^-\right)} \end{aligned}$$

which implies that

$$\begin{aligned} \mathrm{E}_{\mathscr{F}_{n-1}}\left[(p_{n-1}^+ - p_{n-1}^-) A_{n-1}(f_{n-1}) \right] \;\geq\; & -\mathrm{E}\left[\left(p_{n-1}^+ - p_f\right)\left(p_f - p_{n-1}^-\right) \right], \\ \geq \; & -p_f(1-p_f)(1-c_{n-1}) \end{aligned}$$

and the equality, implying $\tilde{p}_n = p_f \; \forall n \geq 1$ and a null global variance, happens only when $\xi_{\boldsymbol{\nu}}/k(\boldsymbol{\nu})$ takes a constant value on the domain $\{\boldsymbol{\nu} \in \mathbb{U}_{n-1},\ \tilde{G}(\boldsymbol{\nu}) \leq 0\}$. The existence of k_{n-1} is needed by having $\mathrm{Supp}(f_{n-1}(\boldsymbol{\nu})) = \mathbb{U}_{n-1}$ and the half-ball condition expresses the lower semicontinuity of the importance distribution function on any point of the failure surface $\mathbb{U}_l$.

Proof (**Proposition 6**)

Observe that $\tilde{p}_{n+1} \geq p_n^-$. Let $\boldsymbol{\nu} \sim f_n$ and assume that $\xi_{\boldsymbol{\nu}} = 1$. Denote $\|.\|$ the L^2 Euclidian norm. Then

$$\begin{aligned} \tilde{p}_{n+1} \;=\; & p_n^- + \exp\left(\frac{\|\boldsymbol{\nu} - \boldsymbol{\nu}_{n,0}\|^2}{2\sigma_n^2} \right) \int_{\mathbb{U}_n} \exp\left(-\frac{\|x - \boldsymbol{\nu}_{n,0}\|^2}{2\sigma_n^2} \right) dx, \\ \leq \; & p_n^- + \exp\left(\frac{\|\boldsymbol{\nu} - \boldsymbol{\nu}_{n,0}\|^2}{2\sigma_n^2} \right) \int_{R^d} \exp\left(-\frac{\|x - \boldsymbol{\nu}_{n,0}\|^2}{2\sigma_n^2} \right) dx, \\ \leq \; & p_n^- + (2\pi)^{d/2} \sigma_n^d \exp\left(\frac{\|\boldsymbol{\nu} - \boldsymbol{\nu}_{n,0}\|^2}{2\sigma_n^2} \right), \\ \leq \; & p_n^- + (2\pi)^{d/2} \sigma_n^d \left[1 + \frac{\|\boldsymbol{\nu} - \boldsymbol{\nu}_{n,0}\|^2}{2\sigma_n^2} (1 + o(1)) \right], \\ \leq \; & p_n^- + (2\pi)^{d/2} D(\sigma_n) + o\left((2\pi)^{d/2} \sigma_n^{d-2} \|\boldsymbol{\nu} - \boldsymbol{\nu}_{n,0}\|^2 \right). \end{aligned}$$

With $\|\boldsymbol{\nu} - \boldsymbol{\nu}_{n,0}\|^2 \leq d$ for all $\boldsymbol{\nu} \in \mathbb{U}$, and since $\eta_n \leq 1 \Rightarrow D(\sigma_n) \leq D(\mu_{n,d}) = (p_n^+ - p_n^-)/(2\pi)^{d/2}$, then

$$\begin{aligned} \tilde{p}_{n+1} \;\leq \; & p_n^- + (p_n^+ - p_n^-) + o\left(d(2\pi)^{d/2} \sigma_n^{d-2} \right), \\ \leq \; & p_n^+ + o\left(n^{-q} \right) \quad \text{from (16)}. \end{aligned}$$

Proof (**Theorem 2**)

Consider the stochastic process defined by $Z_0 = 0$ and

$$Z_n \;=\; \lambda_n \left(\check{p}_{f\,n} - p_f \right) \quad \forall n \geq 1,$$

and

$$\lambda_n \;=\; \prod_{k=1}^{n} (1 - \beta_k)^{-1}.$$

Then from (18) and since $\lambda_n = \lambda_{n+1}(1 - \beta_{n+1})$,

$$\begin{aligned} Z_{n+1} \;=\; & \lambda_n \left(\check{p}_{f\,n} - p_f \right) + p_f \left(\lambda_n - \lambda_{n+1} \right) + \lambda_{n+1} \beta_{n+1} \tilde{p}_{n+1}, \\ =\; & Z_n + \frac{\lambda_n \beta_{n+1}}{1 - \beta_{n+1}} \left[\tilde{p}_{n+1} - p_f \right]. \end{aligned}$$

Because $\tilde{p}_{n+1}$ is unbiased (cf. Lemma 2),

$$\begin{aligned} & \mathrm{E}\left[Z_{n+1} | \mathscr{F}_n \right] \;=\; Z_n \\ & \text{and} \quad \mathrm{E}[Z_n^2] \;=\; \lambda_n^2 V_n^{*WISE} \;<\; \infty \;\text{ at any fixed } n \end{aligned}$$

hence $(Z_n)_n$ is a square integrable martingale. Furthermore, denoting $\Delta Z_n = Z_n - Z_{n-1}$, then

$$\mathrm{E}\left[\Delta Z_{n+1}^2 | \mathscr{F}_n \right] \;=\; (\lambda_{n+1} \beta_{n+1})^2 \, \mathrm{Var}\left[\tilde{p}_{n+1} | \mathscr{F}_n \right]$$

with, according to (28),

$$\begin{aligned} \mathrm{Var}\left[\tilde{p}_{n+1} | \mathscr{F}_n \right] \;=\; & \left(p_n^+ - p_f\right)\left(p_f - p_n^-\right) + \left(p_n^+ - p_n^-\right) A_n(f_n), \\ =\; & p_f(1-p_f)(1-\tilde{d}_n) \end{aligned}$$

with

$$\tilde{d}_n \;=\; \frac{p_n^-}{p_f} + \frac{1-p_n^+}{1-p_f} - \frac{p_n^-(1-p_n^+)}{p_f(1-p_f)} + \frac{b_n(f_n)}{p_f(1-p_f)}.$$

It must be noticed from (9) that

$$1 - \beta_n \;=\; \frac{\sum_{k=1}^{n-1} (1 - d_{k-1})^{-1}}{\sum_{k=1}^{n} (1 - d_{k-1})^{-1}},$$

therefore

$$\lambda_n = \sum_{k=1}^{n}(1-d_{k-1})^{-1}$$

and $\lambda_n\beta_n = (1-d_{n-1})^{-1}$. Thus, defining the increasing (or *angle bracket*) process associated to $(Z_n)_n$ by

$$< Z >_n = \sum_{k=1}^{n} \mathrm{E}\left[\Delta Z_k^2 | \mathscr{F}_{k-1}\right],$$

then

$$\begin{aligned} < Z >_n &= \sum_{k=1}^{n} \lambda_k^2\beta_k^2\left[p_f(1-p_f)(1-c_{k-1}) + B_{k-1}(f_{k-1})\right], \\ &= p_f(1-p_f)\sum_{k=1}^{n}(1-d_{k-1})^{-1}\frac{(1-\tilde{d}_{k-1})}{(1-d_{k-1})}. \end{aligned}$$

Since $\lambda_n^2 V_n^{*WISE} = p_f(1-p_f)\sum_{k=1}^{n}(1-d_{k-1})^{-1}$, then

$$\frac{< Z >_n}{\lambda_n^2 V_n^{*WISE}} = 1+\varepsilon_n$$

with

$$\varepsilon_n = \frac{\sum_{k=1}^{n}(1-d_{k-1})^{-2}a_k}{\sum_{k=1}^{n}(1-d_{k-1})^{-2}}$$

and $a_k = d_{k-1} - \tilde{d}_{k-1}$. Under assumptions (i) to (iii), $a_n \xrightarrow{I\!P} 0$. Then $\exists N \in I\!N$ such that $\forall n \geq N$, $|a_n| \leq \sum_{k=1}^{n-1}|a_k|$. Indeed, assuming the contradiction of this assumption implies that $\forall N$, $\exists n \geq N$ such that any partial sum of the $|a_k|$ with $k < n$ be upperly bounded by $|a_n|$; this would imply that the sum of term $|a_n|$ converges toward 0, which is only possible if all a_n are zero. Finally, since $(1-d_{k-1})^{-2} > 1$ for $k > 1$, then $\forall n \geq N$, $|a_{n+1}| < \sum_{k=1}^{n-1}(1-d_{k-1})^{-2}|a_k|$. Equivalently, $\forall n \geq N$, $\frac{\varepsilon'_{n+1}}{\varepsilon'_n} < 1$ where

$$\varepsilon'_n = \sum_{k=1}^{n}(1-d_{k-1})^{-2}|a_k| \Big/ \sum_{k=1}^{n}(1-d_{k-1})^{-2}.$$

The sum of general term ε'_n also converges, then necessarily $\varepsilon'_n \to 0$ and since $|\varepsilon_n| \leq \varepsilon'_n$, the sequence $\{\varepsilon_n\}$ converges absolutely toward 0. Then

$$\frac{< Z >_n}{\lambda_n^2 V_n^{*WISE}} \xrightarrow[n\to\infty]{a.s.,I\!P} 1. \tag{29}$$

For all $1 \leq k \leq n$, denote now $H_{n,k} = \lambda_n\sqrt{V_n^{*WISE}}(1-d_{k-1})$ which tends toward infinity when $n \to \infty$, even though $k = n$. Note that $\forall \varepsilon > 0$,

$$\begin{aligned} &P\left(\left|\frac{|\Delta Z_k|}{\lambda_n\sqrt{V_n^{*WISE}}}\right| < \varepsilon\right) \\ &\quad = P\left(p_f - p_{k-1}^- - \varepsilon H_{n,k} < \frac{\delta_{\boldsymbol{\nu}}}{f_{k-1}(\boldsymbol{\nu})} < p_f - p_{k-1}^- + \varepsilon H_{n,k}\right) \end{aligned}$$

so that this probability tends to $P(-\infty < \delta_{\boldsymbol{\nu}} < \infty) = 1$. Consequently, $\forall \varepsilon > 0$, $\mathbb{1}_{\{|\Delta Z_k| \geq \varepsilon\sqrt{\lambda_n^2 V_n^{*WISE}}\}} \xrightarrow{I\!P} 0$ and the sum $\sum_{k=1}^{n}\mathrm{E}[\Delta Z_k^2 \mathbb{1}_{\{|\Delta Z_k| \geq \varepsilon\sqrt{\lambda_n^2 V_n^{*WISE}}\}} | \mathscr{F}_k]$ is bounded, so a Lindeberg condition is satisfied:

$$\frac{1}{\lambda_n^2 V_n^{*WISE}}\sum_{k=1}^{n}\mathrm{E}\left[\Delta Z_k^2 \mathbb{1}_{\{|\Delta Z_k| \geq \varepsilon\sqrt{\lambda_n^2 V_n^{*WISE}}\}} | \mathscr{F}_k\right] \xrightarrow{I\!P} 0. \tag{30}$$

Finally, (29) and (30) prove the martingale central limit theorem (Bercu 2008):

$$\frac{1}{\sqrt{\lambda_n^2 V_n^{*WISE}}}\lambda_n\left(\breve{p}_{f\,n} - p_f\right) \xrightarrow{\mathscr{L}} \mathscr{N}(0,1).$$

Finally, from the fact that $\lim_{n\to\infty} < Z >_n = \infty$ and using Theorem 3.6 in Bercu (2008),

$$\frac{\lambda_n\left(\breve{p}_{f\,n} - p_f\right)}{< Z >_n} \xrightarrow{a.s.} 0.$$

Using (29) and Slutsky's theorem, one has $(\lambda_n V_n^{*WISE})^{-1}(\breve{p}_{f\,n} - p_f) \xrightarrow{a.s.} 0$ with $\lambda_n V_n^{*WISE} = p_f(1-p_f)$ $\forall n \geq 1$, which proves the strong consistency of $\breve{p}_{f\,n}$.

Proof **(Proposition 7)**

The proof is simply based on the Azuma-Hoeffding inequality applied to the martingale $(Z_n)_n$ defined in the Proof of Theorem 2. Indeed, since

$$a_n \leq \Delta Z_n \leq a_n + (1-d_{n-1})^{-1}/f_{n-1}(\boldsymbol{\nu}_n)$$

with $a_n = (1-d_{n-1})^{-1}(p_{n-1}^- - p_f)$, then $\forall \varepsilon' > 0$,

$$P\left(Z_n \leq -\varepsilon'\right) \leq \exp\left(-\frac{2\varepsilon'^2}{\sum_{k=1}^{n}(1-d_{k-1})^2/f_{k-1}^2(\boldsymbol{\nu}_k)}\right)$$

The result can then be deduced using $\varepsilon = \varepsilon'/(\lambda_n p_f)$.

Electronic Supplementary Material

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 [6]. The d-dimensional volume is calculated recursively by exploring all $n-1$-dimensional "slices" of the d-th dimension. See [8,1,3] 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 [4,2]. Non-optimized pseudo-code follows to be used for direct implementation.

Let Δ_n be the $n \times d$ matrix of n vertexes $(\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n)$ defining the union of hypercubes (for an example, see Figure 11). 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 $\Delta_n^{(i)} = \{\boldsymbol{v}'_i, \ldots, \boldsymbol{v}'_n \in \Delta'_n\}$.
   (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 $\widetilde{\texttt{Vol}}_n^{(i)} = \max\{v \in \Delta_n^{(i)}\}$;
   - 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 $\texttt{Vol}_n = \texttt{Vol}_n + \texttt{Vol}_n^{(i)}$.

To our knowledge, this algorithm remains little used for dimension d larger than 2 or 3. This is not surprising since it losses its intuitive aspects and because its complexity $C(n,d)$ (number of runs for a d-dimensional hypervolume between n points) is $O(n^d)$. This appears when looking at the first developments of $C(n,d)$:

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

Note however than the fastest currently known version of this algorithm has been proposed by Overmars and Yap [7], which runs in time $O(n^{d/2} \log n)$ for $d \geq 3$. An alternative approach has been proposed by Chlebus [3] with the same asymptotic performance, although its exposition has been 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 been recenlty proposed by Chan [2]. We suggest that some ideas of possible improvements could come from a parallel with multi-objective optimization contexts (cf. Remark 1 in the article). Indeed, algorithms running in polynomial time $n^{k_1} d^{k_2})$ to compute hypervolume metrics of Pareto frontiers have already been proposed by Fleischer (2003).

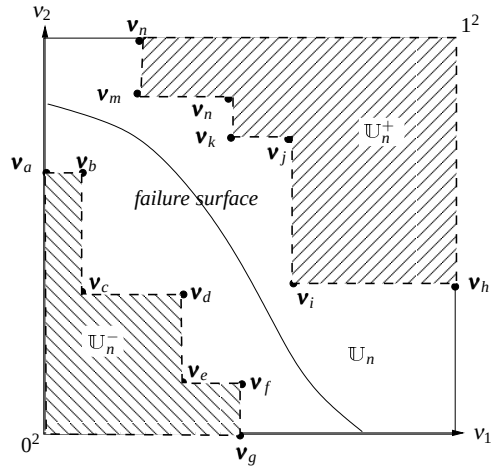

Fig. 11 Two-dimensional uniform space after $n = 14$ MRM iterations. Points $\Delta_{14} = \{0^2, \boldsymbol{v}_a, \boldsymbol{v}_b, \boldsymbol{v}_c, \boldsymbol{v}_d, \boldsymbol{v}_e, \boldsymbol{v}_f, \boldsymbol{v}_g\}$ have positive signatures and are vertexes of $\mathbb{U}_n^-$. Points $\{\boldsymbol{v}_h, \boldsymbol{v}_i, \boldsymbol{v}_j, \boldsymbol{v}_k, \boldsymbol{v}_l, \boldsymbol{v}_m, \boldsymbol{v}_n, 1^2\}$ have zero signatures and are vertexes of $\mathbb{U}_n^+$.

References

1. de Berg, M., van Kreveld, M., Overmars, M., Schwarzkopf, O., Computational Geometry Algorithms and Applications. Springer-Verlag (1997).
2. Chan, T.M., A (slightly) faster algorithm for Klee's measure problem, In: Proceedings of the twenty-fourth annual symposium on Computational Geometr, College Park, MD, USA, 94-100 (2008).
3. Chlebus, B.S., On the Klee's measure problem in small dimensions, In: Proceedings of the 25th Conference on Current Trends in Theory and Practice of Informatics (SOFSEM-98), Slovakia (1998).
4. Erickson, J., Klee's measure problem, Open Problems, University of Illinois at Urbana-Champaign, URL: *http://theory.cs.uiuc.edu/ jeffe/open/klee.html* (1998).
5. Fleischer, M., The measure of Pareto optima. Applications to multi-objective metaheuristics, In: Fonseca, C.M., Fleming, P.J., Zitzler, E., Deb, K., Thiele, L. (eds): Evolutionary multi-criterion optimization. Second International Conference EMO 2003, Faro, Portugal, Springer. Lecture Notes in Computer Sciences, 262, 519-533 (2003).
6. van Leeuwen, J., Wood, D., The measure problem for rectangular ranges in $d-$space, Journal of Algorithms, 2, 282-300 (1981).

7. Overmars, M. H., Yap, C.-K., New upper bounds in Klee's measure problem, SIAM Journal of Computing 20(6), 1034-1045 (1991).
8. Shamos, M. I., Hoey, D., Geometric intersection problems, Proc. 17th IEEE Symp. Foundations of Computer Science (FOCS '76), 208-215 (1976).