Une première exploration de l'utilisation de l'hypothèse de monotonie pour l'évaluation de probabilité de dépassement de seuil

[work in progress]

N. Bousquet

based on works by **E. de Rocquigny** (ECP) and **Ph. Limbourg** (Université Duisburg-Essen)

EDF R& D

Context

We are interested in predicting the behavior of a unidimensional variable

$$Z = G(\mathbf{X})$$

where

- G is a deterministic function (computer code)
- $\mathbf{X} = (X_1, \ldots, X_d)$ is a $d-$dimensional random vector (stationary conditions)

More precisely, given a fixed margin z^*, we want to estimate accurately $p = P(Z \leq z^*)$

Some examples of EDF contexts:

- G = hydraulical code, Z = water level, z^* = dyke height, $\mathbf{X}$ = geography
- $G(\mathbf{X}) = \mathbf{R} - \mathbf{S}$ (nuclear reactor core), R = resistance, S = solicitation, $z^* = 0$

Without loss of generality, we will always consider $z^* = 0$ in the following

Constraints

Statistical/probabilist methods to estimate p usually use a *Design of Experiments* (DOE), ie. a set of values $\mathbf{x_1}, \ldots, \mathbf{x_n}$ one propagates through G

Such methods should take account of the following items:

1. **Non-intrusiveness**: G is most often seen as a (partial) *black box*
2. **Parsimony**: a run of G can be very time-consuming $\Rightarrow$ fixed number of runs of G independently of p
3. **Conservatism**: one should avoid to underestimate p

Expected properties of a good estimator $\hat{p}_n$ of p:

1. **Null or slightly positive bias** (conservatism)
2. **Consistency** : $\hat{p}_n \xrightarrow{a.s.} p$
3. **Weak variance** (e.g., with respect to a classic Monte Carlo estimator)
4. **Robustness** : $\lim_{p \to 0} \mathrm{CV}\left[\hat{p}_n\right] < \infty$ [Rubino and Tuffin 2009]
5. **Parsimony in theoretical conditions**: regularity of G, validity of linearization, curse of dimensionality ....

Monotonous computer codes G

In EDF contexts, one can often find a parametrization of X such that G is continuously *increasing* (monotonous) w.r.t. X: for any $\delta > 0$,

$$G\left(x_1, \ldots, x_{i-1}, x_i + \delta, x_{i+1}, \ldots, x_d\right) \quad \leq \quad G\left(x_1, \ldots, x_{i-1}, x_i, x_{i+1}, \ldots, x_d\right).$$

Makes sense in many areas (structural safety assumption)

This assumption could make locally non-monotonous models be more conservative

---

Remark

- Let $\hat{G}(\mathbf{X}, \epsilon) = G(\mathbf{X})(1 + \epsilon)$ with $\epsilon =$ *model uncertainty + observational errors*
- Assume ϵ can be calibrated using noisy data $Z = \hat{G}(\mathbf{X}, \epsilon)$ and a simplified analytic model $\hat{G}^*$
- $\hat{G}$ remains monotonous (increasing)

Main points

1. Using the monotonicity of G w.r.t. X to bound p
   - providing conservative (biased) estimators of p
   - independent X_i
   - dependent X_i
2. Estimating p without supplementary runs of G
   - Consistency
   - Robustness
   - Fastness
   - Sensitivity studies
3. Simulation studies (generic case-study)
4. Connections with other problems in mathematics

Some useful notations

Denote $(\boldsymbol{D}, \mathcal{A}, P)$ the probability space of stochastic inputs $\mathbf{X}$, with dim $\boldsymbol{D} = d$

Denote $(f_{\mathbf{X}}, F_{\mathbf{X}})$ the pdf and cdf of $\mathbf{X}$ with support $\boldsymbol{D}$

Any $\mathbf{x} \in \boldsymbol{D}$ such that $G(\mathbf{x}) \leq 0$ (resp., $G(\mathbf{x}) > 0$) will be called a *failure point* (resp., a *safety point*)

Denote $\hat{p}_n$ any estimator of

$$p \quad = \quad P\left(G(\mathbf{X}) \leq 0 | \mathbf{X} \in \boldsymbol{D}\right) \quad = \quad \int \mathbb{1}_{\{G(\mathbf{x}) \leq 0\}} f_{\mathbf{X}}(\mathbf{x}) \, d\mathbf{x}$$

based on n calls to G

A generic case-study

In the numerical illustrations to come, denote:

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

where

$$X_i \quad \sim \quad \mathcal{G}(i+1, 1)$$

Thus

$$Y_d \quad \sim \quad \mathcal{B}_e \left(2, \frac{(d+1)(d+2)}{2} - 3 \right)$$

Denote $q_{d,\alpha}$ the $\alpha-$order quantile of Y_d
Thus we will associate the code

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

to the known exceedance probability α

Using code monotony to bound p

Note $\mathbf{X} \succeq \mathbf{X}' \Leftrightarrow X_i \geq X_i'$ for $i = 1, \ldots, d = \dim(\mathbf{X})$

- If $\mathbf{x} \succeq \mathbf{x}'$ and $G(\mathbf{x}) \leq 0$, then $G(\mathbf{x}') \leq 0$ (*failure points* $(\mathbf{x}, \mathbf{x}')$)
- If $\mathbf{y} \preceq \mathbf{y}'$ and $G(\mathbf{y}) > 0$, then $G(\mathbf{y}') > 0$ (*safety points* $(\mathbf{y}, \mathbf{y}')$)

Lemma

Denote $\boldsymbol{D}_{\mathbf{x}}^{-} = \{\mathbf{x}' \in \boldsymbol{D} \mid \mathbf{x}' \preceq \mathbf{x}\}$ and $\boldsymbol{D}_{\mathbf{y}}^{+} = \{\mathbf{y}' \in \boldsymbol{D} \mid \mathbf{y}' \succeq \mathbf{y}\}$. Then:

$$p^{-} = P(\mathbf{X} \in \boldsymbol{D}_{\mathbf{x}}^{-}) \leq p \leq p^{+} = 1 - P(\mathbf{X} \in \boldsymbol{D}_{\mathbf{y}}^{+})$$

One can imagine iterative algorithms to decrease the width of $[p^{-}, p^{+}]$ and provide conservative estimates of p

MRM = Monotonous Reliability Method: principles and heuristics recently proposed by de Rocquigny (2009)

Space transformation

Often a preliminary task in propagation methods used in structural reliability methods (ex: FORM/SORM)

Simplify the understanding / computation of the bounds

From $\boldsymbol{D}$ to hypercube $\mathbb{U} = [0,1]^d$

Working with $\tilde{G}(\boldsymbol{\nu}) = G \circ \Psi_{F_{\mathbf{X}}}^{-1}(\boldsymbol{\nu}) \;=\; G(\mathbf{X})$ where $\boldsymbol{\nu} = (\mu_1, \ldots, \mu_d) \overset{\text{iid}}{\sim} \mathcal{U}_{\mathbb{U}}$

Independent (continuous) X_i: distribution functions

Dependent (continuous) X_i: Rosenblatt / Nataf transforms

- Some conditions must be respected on the copula or the conditionning to preserve monotony

Visual effects of monotony in the $\mathbb{U}$ hypercube

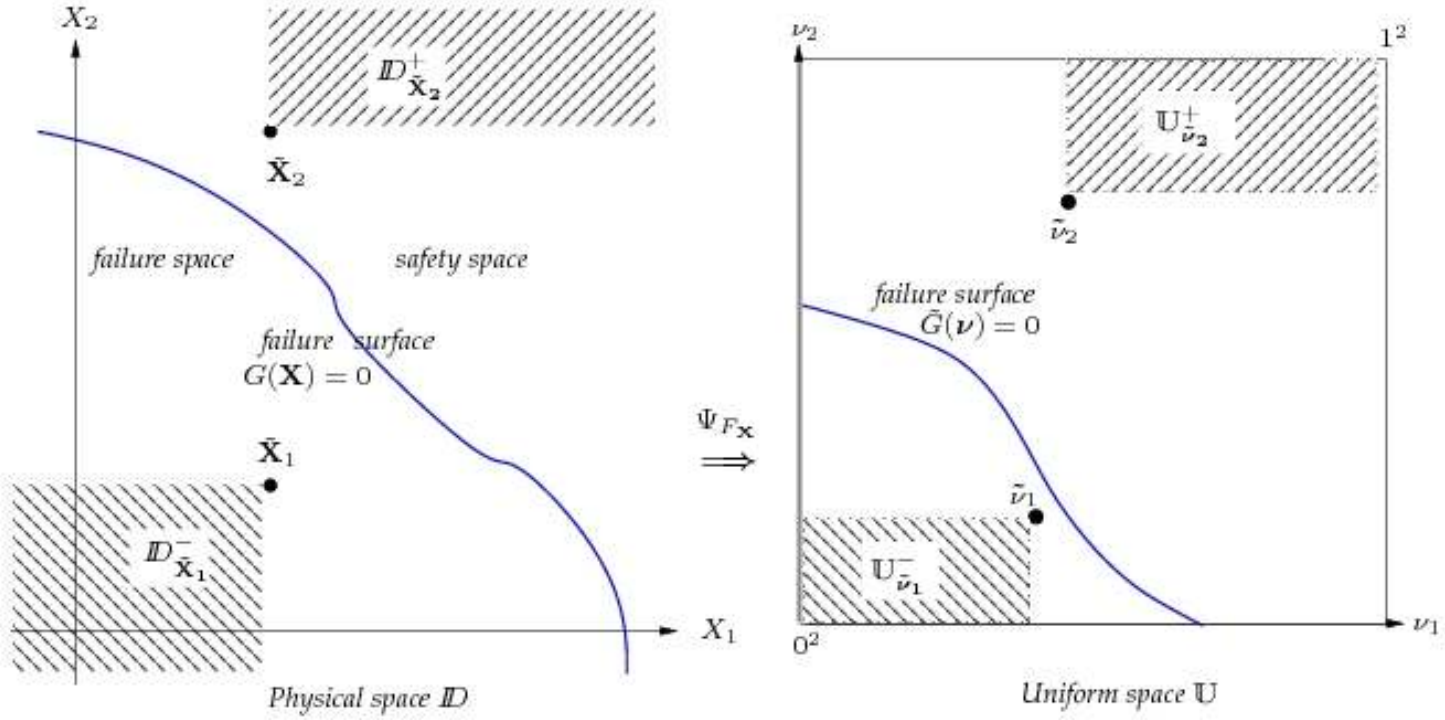


Figure 1: A two-dimensional case ($d = 2$). Axes define canonical bases. Transforming physical space $\mathbb{D}$ in uniform space $\mathbb{U}$ simplifies the hypervolumic calculus of

$$P(\mathbf{X} \in \mathbb{D}^-_{\tilde{\mathbf{X}}_1}) = P(\boldsymbol{\nu} \in \mathbb{U}^-_{\tilde{\boldsymbol{\nu}}_1}) = 1 - \prod_{i=1}^{d} \tilde{\boldsymbol{\nu}}_{1,i} \quad \text{and}$$

$$P(\mathbf{X} \in \mathbb{D}^+_{\tilde{\mathbf{X}}_2}) = P(\boldsymbol{\nu} \in \mathbb{U}^+_{\tilde{\boldsymbol{\nu}}_2}) = \prod_{i=1}^{d} \tilde{\boldsymbol{\nu}}_{2,i}$$

Sequential DOE (one ν - one modification of one bound)

We call $\xi_{\nu} = \mathbb{1}_{\{\tilde{G}(\nu) \leq 0\}}$ the *signature* of ν

Progressive bounds $(p_n^-, p_n^+) = (\text{Vol}(\mathbb{U}_n^-), 1 - \text{Vol}(\mathbb{U}_n^+))$

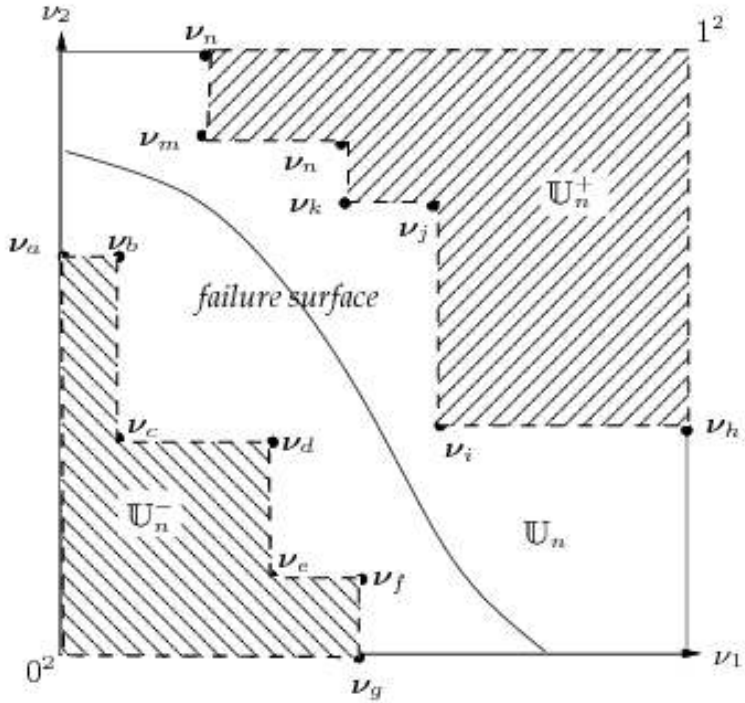


Figure 2: Two-dimensional uniform space after $n = 14$ MR iterations. Points $\{0^2, \nu_a, \nu_b, \nu_c, \nu_d, \nu_e, \nu_f, \nu_g\}$ have positive signatures and are vertexes of $\mathbb{U}_n^-$. Points $\{\nu_h, \nu_i, \nu_j, \nu_k, \nu_l, \nu_m, \nu_n, 1^2\}$ have zero signatures and are vertexes of $\mathbb{U}_n^+$.

Computing the volumes (= computing the bounds): 2 possibilities

Approx. Monte Carlo computation.

$$\mathrm{Vol}(\mathbb{U}_{\boldsymbol{n}}^{-}) \quad = \quad \int_{\mathbb{U}_{\boldsymbol{n}}^{-}} f_{\mathbb{U}}(\boldsymbol{\nu})\, d\boldsymbol{\nu} \quad = \quad \lim_{M\to\infty} \frac{1}{M} \sum_{i=1}^{M} \mathbb{1}_{\{\boldsymbol{\nu_i}\in\mathbb{U}_{\boldsymbol{n}}^{-} \,|\, \boldsymbol{\nu_i}\in\mathbb{U}\}},$$

We "loose" the gain of 100%-confidence bounds...

Exact computation. Volume of unions of hyperrectangles = Klee's measure problem (computational geometry)

- Advantage to work in a uniform space
- Recursive sweepline algorithm (*balayage*) [Chlebus 1998, Chan 2008]
- Cost $\mathcal{O}\left(n^{d/2} \log n\right)$ in dimension d (n = number of vertexes) must be compared to the cost of G

We suggest to switch from a sweepline algorithm to Monte Carlo in high dimensions

Initialization

Using a deterministic DOE

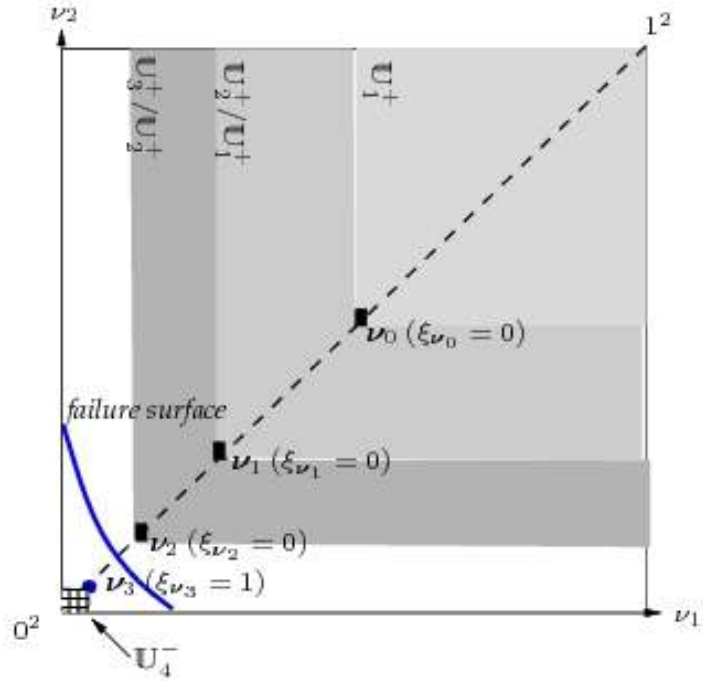
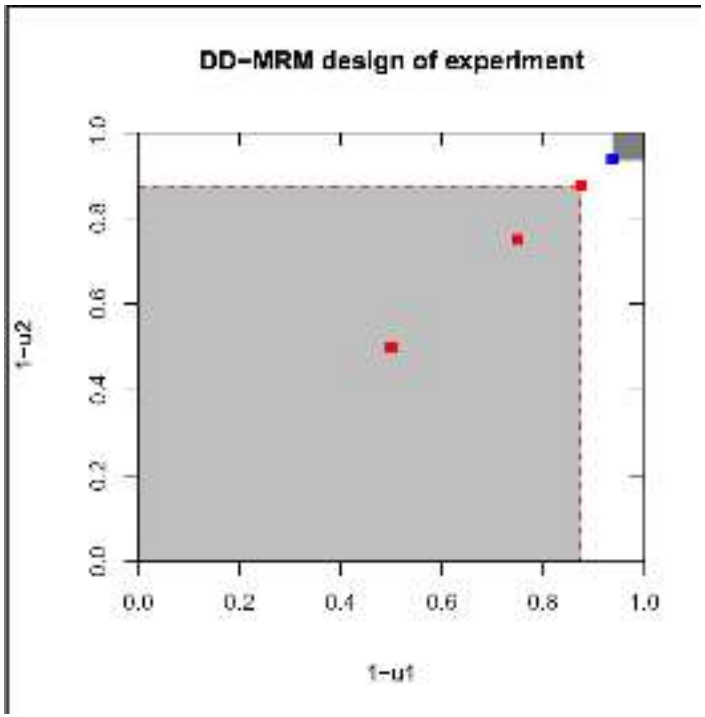


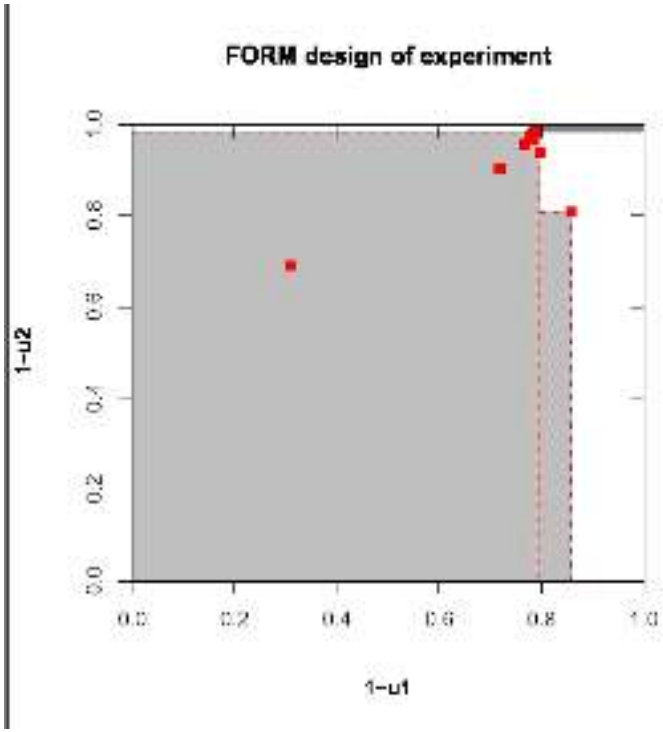
Figure 3: Diagonal deterministic (DD-MRM) strategy, assuming a very low p_f, stopping after 4 steps.

Other possibility: FORM/SORM algorithms (structural reliability)

Give a crude **design point** after some iterations, close to the failure surface

Case-study 2D ($\alpha = 5\%$): DD-MRM and FORM-MRM Designs of Experiments





Stopping criterion

Relative *length of certain interval*:

$$\mathrm{LCI}_n(p) \quad = \quad (p_n^+ - p_n^-)/p$$

Define

$$\rho_n \quad = \quad \mathrm{LCI}_n(p_n^-) \quad = \quad p_n^+/p_n^- - 1$$

more practical since, $\forall \epsilon > 0$ (typic. $\epsilon \sim 5\%$)

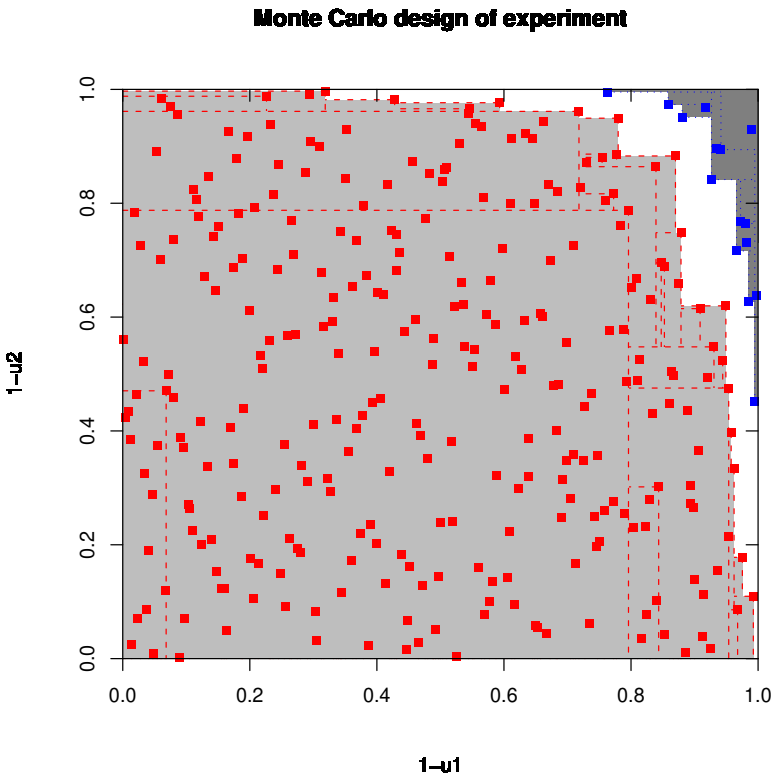
$$\rho_n \leq \epsilon \quad \Rightarrow \quad \mathrm{LCI}_n(p) \leq \epsilon$$

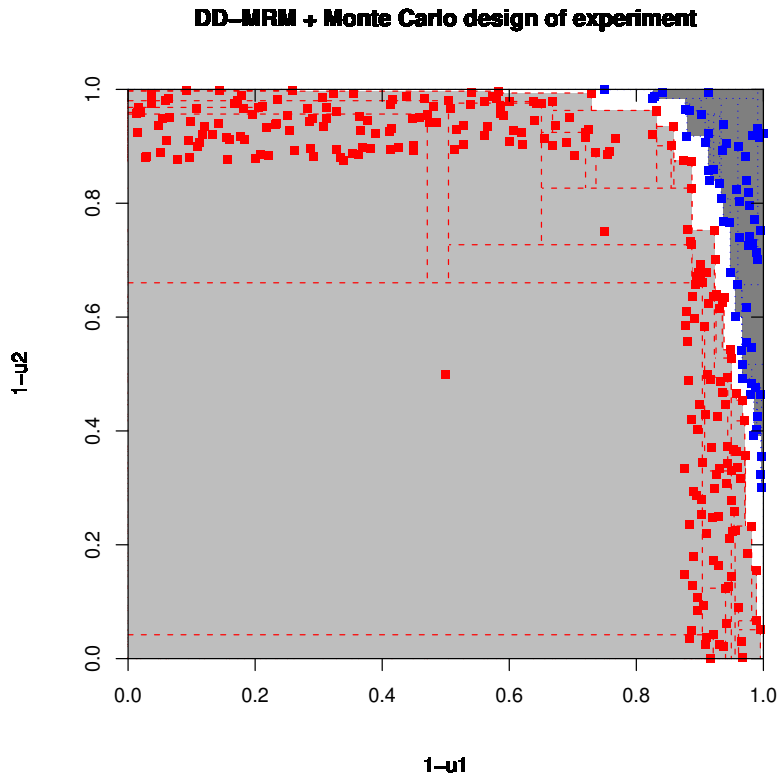
Stopping at iteration $k_0 \leq 1$ after getting non-trivial bounds $(p_n^-, p_n^+) \notin \{0, 1\}$

$\Rightarrow \rho_{k_0} < \infty$

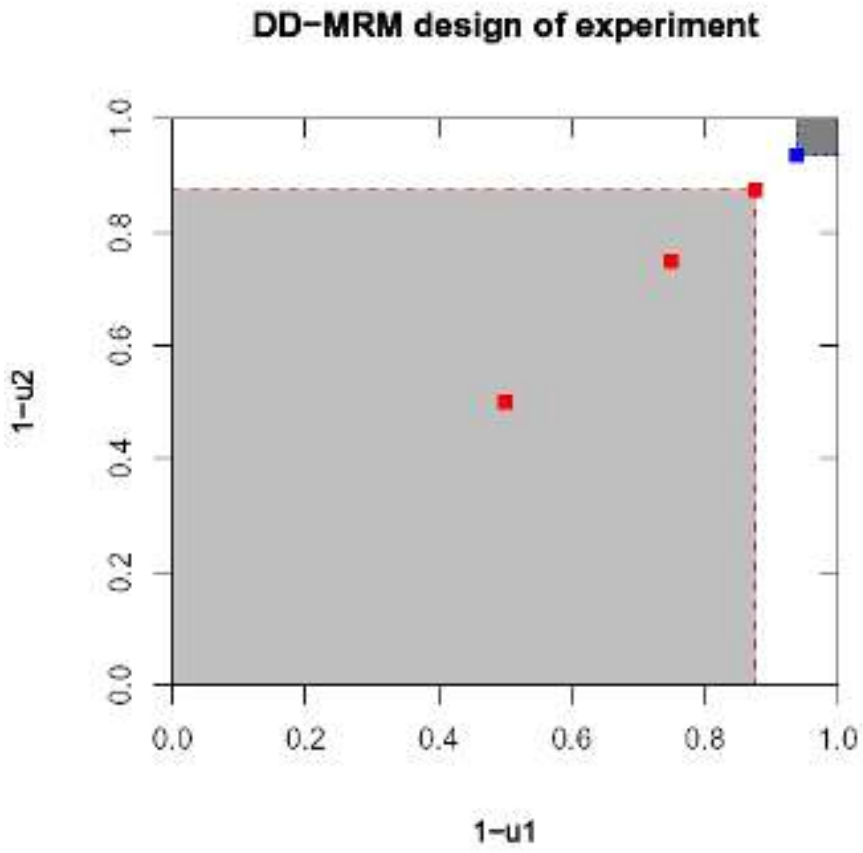
$$\log \sum_{k=1}^{k_0} 2^{-k} \quad \leq \quad d \log(1-p)$$

Illustration





Illustration



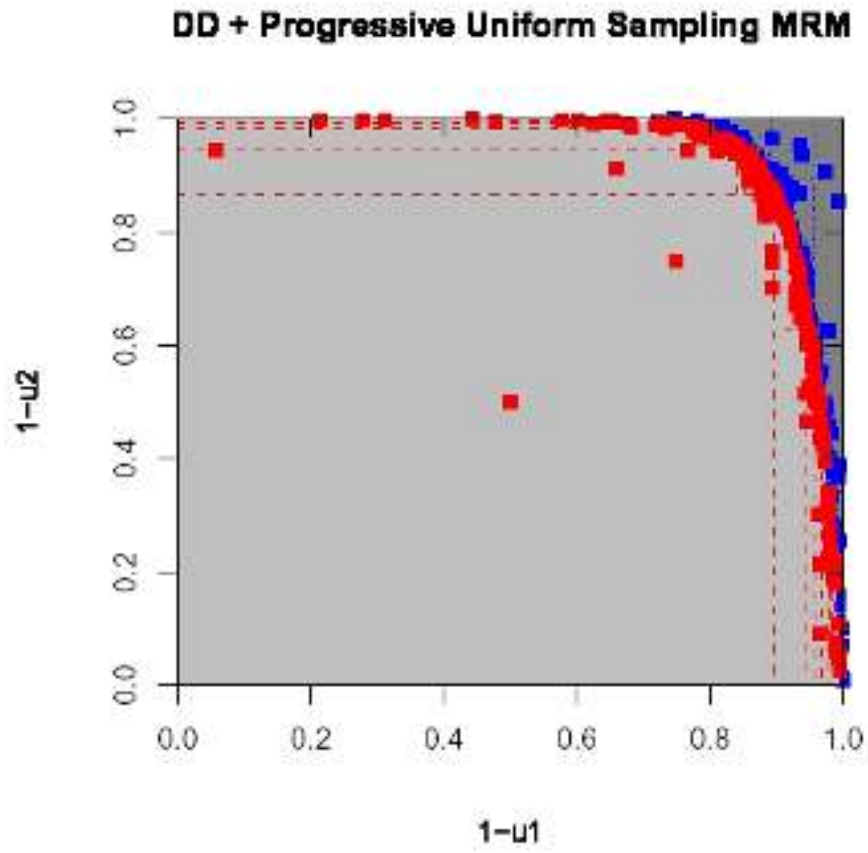
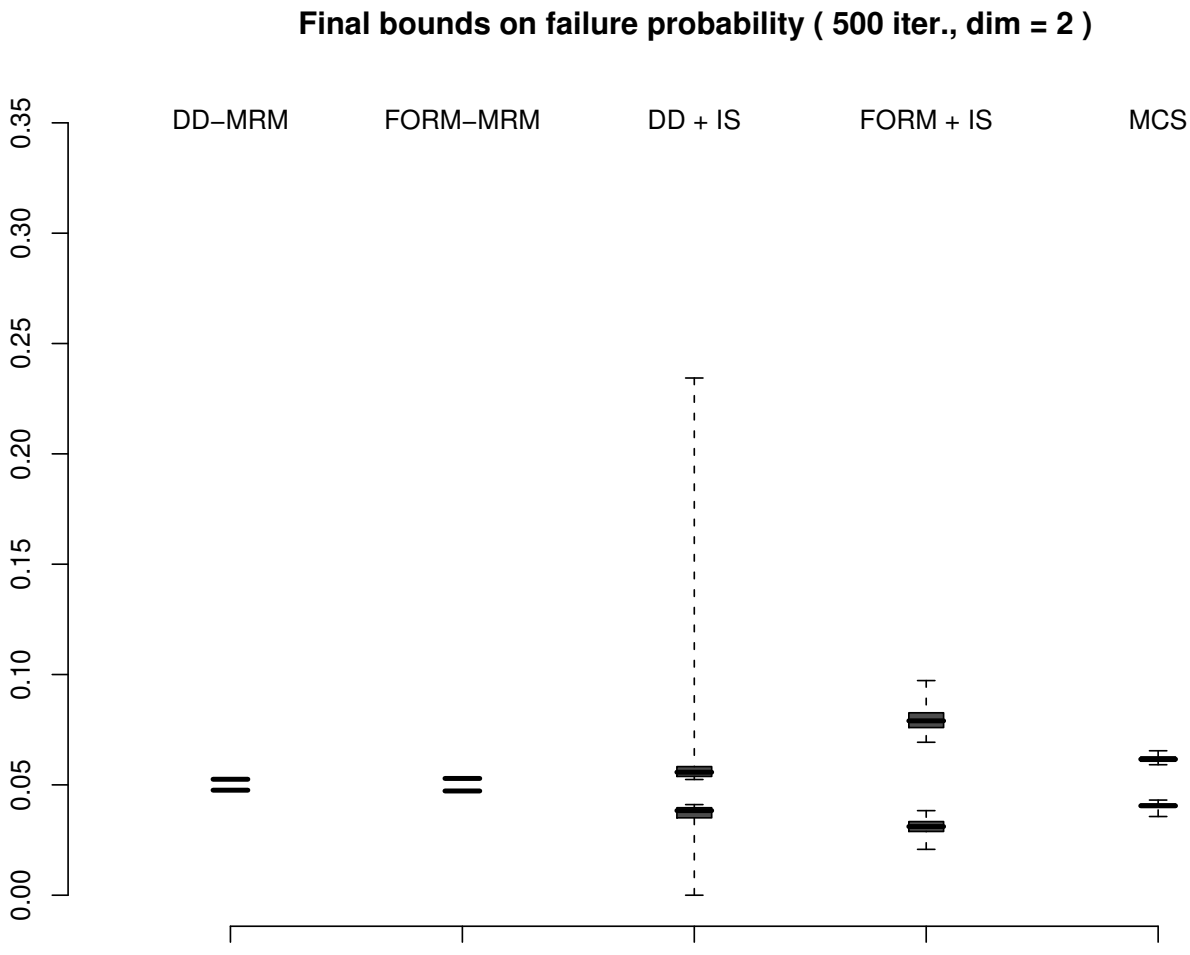


Illustration: computing final bounds on p



Sensitivity studies of bounds without new runs of G

Sensitivity study need modifying F_X in $\tilde{F}_X$ in a variational class

Recomputing bounds without supplementary calls to G

- Physical spaces $(\Psi_{F_X}^{-1}(\mathbb{U}_n^-), \Psi_{F_X}^{-1}(\mathbb{U}_n^+))$ remain dominated whatever the choice made on input distributions
- Reflect the properties of the deterministic function G, independently of the likelihood of each sample point in $\boldsymbol{D}$.

$$
\begin{aligned}
p_n^-(\epsilon) &= \int_{\Psi_{F_X}^{-1}(\mathbb{U}_n^-)} f_X^\epsilon(\mathbf{x})\, d\mathbf{x}, \\
p_n^+(\epsilon) &= 1 - \int_{\Psi_{F_X}^{-1}(\mathbb{U}_n^+)} f_X^\epsilon(\mathbf{x})\, d\mathbf{x}.
\end{aligned}
$$

Could lead to a supplementar definition of what is a robust estimation of p is a perturbed estimator stays between the same bounds

Stochastic Designs of Experiments (DOE) and statistical estimation

Generating bounds for p can be too crude

Explorating a large space $\Rightarrow$ too conservative upper bound (without real interest)

A more precise statistical estimator could be possibly yielded using a progressive stochastic DOE

Restart the notation $(\mathbb{U}_0^+, \mathbb{U}_0^-, p_0^+, p_0^-)$ after $N-1$ introductive deterministic steps with $N \geq k_0 + 1$

Assume now at each step i, $\boldsymbol{\nu}_i \sim f_{i-1}$ is sampled stochastically on $\mathbb{U}_{i-1} = \mathbb{U}/(\mathbb{U}_{i-1}^- \cup \mathbb{U}_{i-1}^+)$

Convergence of bounds and criteria

Assume $\lim_{n\to\infty} \bigcup_{i=1}^{n} \mathrm{Supp}\,(f_i) \equiv \mathbb{U}_0$. Then

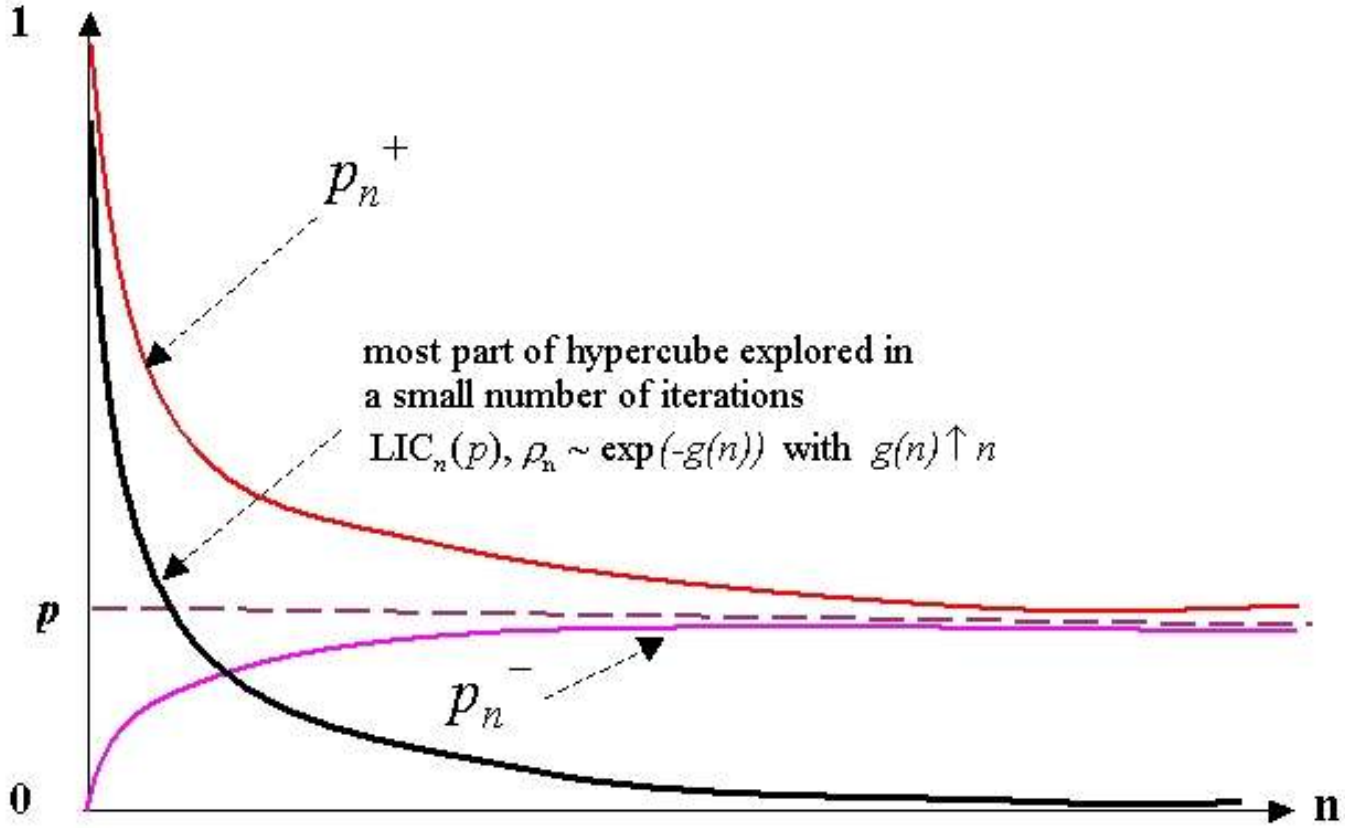
$$\begin{aligned} p_n^-, p_n^+ &\xrightarrow{a.s.} p, \\ \rho_n, \mathrm{LCI}_n(p) &\xrightarrow{a.s.} 0 \quad \text{(strict decreasing)} \end{aligned}$$

Any $n-$estimator of p located in $[p_n^-, p_n^+]$ is consistent

---

One can expect $(\rho_n, \mathrm{LCI}_n(p))$ exponentially decrease

Illustration



Definition and study of some statistical estimators

---

1. Progressive uniform sampling in non-dominated subspaces (no information on the failure surface)
2. Using any arbitrary progressive importance sampling (assuming information on the failure surface)

How we can do better than the reference estimator: Monte Carlo standard?

Reminder: Monte Carlo standard (MCS)

$$\hat{p}_n = \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{\{G(\mathsf{x}_i)\leq 0\}} \quad \text{with } X_i \overset{iid}{\sim} f_{\mathsf{X}}$$

$$\mathrm{Var}(\hat{p}_n) = \mathrm{V}_n^{MC} = \frac{p(1-p)}{n}$$

Advantages:

1. best adapted to “blind" cases (no requirements on G)
2. consistency/asymptotic normality of $\hat{p}_n$

Defects:

1. **no parsimony:** needs 10^{q+2} runs of G for a good estimation of 10^{-q}; unrealistic approach
2. **lack of robustness:** highly unreliable when $p \to 0$

$$\lim_{p\to 0} \mathrm{CV}\left[\hat{p}_n\right] = \infty$$

Uniform importance sampling

At each step i, sample $\boldsymbol{\nu}_i$ uniformly in $\mathbb{U}_{i-1}$

The occurence of a nonzero signature $\xi_{\boldsymbol{\nu}_i}$ follows a Bernoulli distribution with probability parameter

$$\gamma_i = P\left(\tilde{G}(\boldsymbol{\nu}) \leq 0 \middle| \boldsymbol{\nu} \in \mathbb{U}_{i-1}\right)$$

Using Bayes' rule, one has

$$\gamma_i = \frac{P\left(\tilde{G}(\boldsymbol{\nu}) \leq 0 | \boldsymbol{\nu} \in \mathbb{U}\right) - P\left(\tilde{G}(\boldsymbol{\nu}) \leq 0 | \boldsymbol{\nu} \in \mathbb{U}/\mathbb{U}_{i-1}\right) P\left(\boldsymbol{\nu} \in \mathbb{U}_{i-1}\right)}{P\left(\boldsymbol{\nu} \in \mathbb{U}_{i-1}\right)}$$

$$= \frac{p - p_{i-1}^{-}}{p_{i-1}^{+} - p_{i-1}^{-}}$$

Data-dependent maximum likelihood estimator (MLE)

Likelihood of dependent data

The joint density of n successive signatures is

$$L_n(p) = \prod_{i=1}^{n} \left(\frac{p - p_{i-1}^-}{p_{f,i-1}^+ - p_{i-1}^-} \right)^{\xi_{\nu_i}} \left(\frac{p_{f,i-1}^+ - p}{p_{f,i-1}^+ - p_{i-1}^-} \right)^{1-\xi_{\nu_i}}$$

Existence, unicity, consistency

There exists a unique MLE $\hat{p}_n \in (p_{n-1}^-, \tilde{p}_{n-1}^+) \in [p_{n-1}^-, p_{n-1}^+]$, with $\hat{p}_n \xrightarrow{a.s.} p$ and

$$\begin{aligned} \hat{p}_n &= \sum_{k=1}^{n} \tilde{\omega}_k(\hat{p}_n)\, p_k \Big/ \sum_{k=1}^{n} \tilde{\omega}_k(\hat{p}_n) \\ \text{with } \bar{p}_k &= p_{k-1}^- + \left(p_{k-1}^+ - p_{k-1}^-\right) \xi_{\nu_k} \quad \text{(unbiased local estimator)} \\ \tilde{\omega}_k(x) &= \left[\left(x - p_{k-1}^-\right)\left(p_{k-1}^+ - x\right)\right]^{-1} \quad \text{(increasing weights)} \end{aligned}$$

Illustration

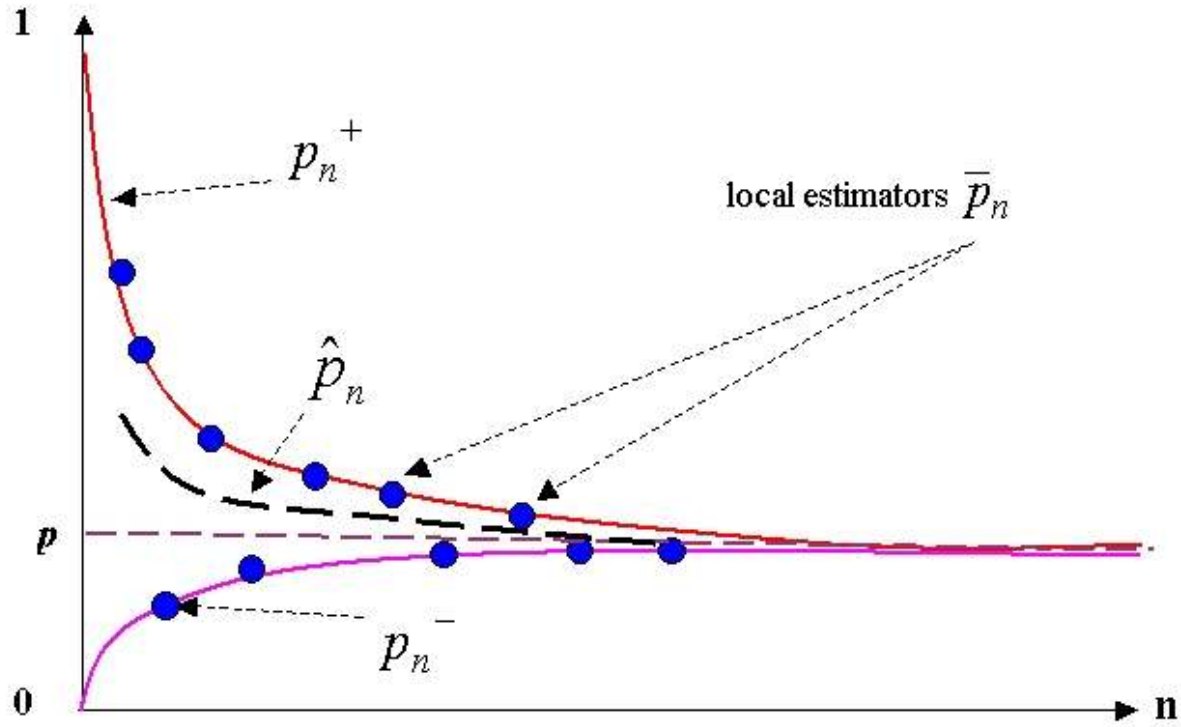
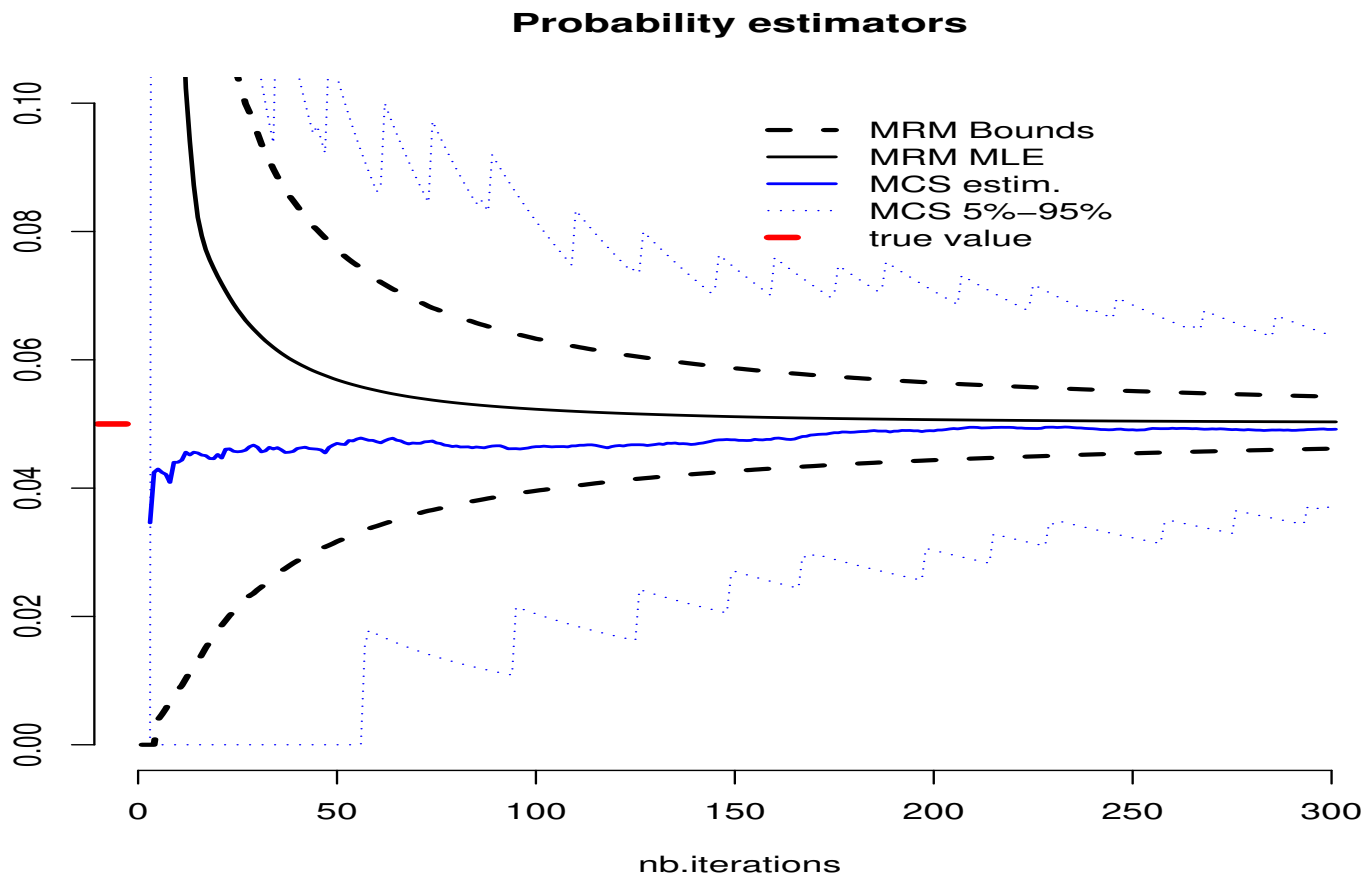


Illustration: $d = 2$, $\alpha = 5.10^{-2}$ (averaged on 100 MRM parallel runs)



Behavior of $\hat{p}_n$

Denote $\mathcal{F}_n$ the $\sigma-$algebra generated by n successive MRM steps

Denote $\ell_n(p) = \log L_n(p)$

Usual regularity conditions

If the failure surface $\{\boldsymbol{\nu} \in \mathbb{U},\ \tilde{G}(\boldsymbol{\nu}) = 0\}$ is not piecewise constant, then

$$\begin{aligned} \mathrm{E}_{\mathcal{F}_n}\left[\ell_n'(p)\right] &= 0 \\ \mathrm{Var}_{\mathcal{F}_n}\left[\ell'(p)\right] &= \Lambda_n = \sum_{i=1}^{n} \mathrm{E}_{\mathcal{F}_i}\left[\tilde{\omega}_i(p)\right] < \infty \quad \text{(Fisher information)} \end{aligned}$$

else p can be exactly reached in a finite number of steps

Asymptotic behavior of $\hat{p}_n$?

- Asymptotic normality of $\ell_n'(p)$ / Law of Large Numbers (LLN) for $\ell_n''(\hat{p}_n)$

using the fact that $\ell_n'(p)$ is a martingale [Crowder 1983, Heijmans and Magnus 1986]

Asymptotism

Hypothesis A: distances between bounds (p_n^-, p_n^+) and p become indistinguisable of their average behavior ($\sim$ Lindeberg condition)

$$\sum_{i=1}^{n} \left(\tilde{\omega}_i - \mathsf{E}_{\mathcal{F}_i}\left[\tilde{\omega}_i\right]\right) \quad = \quad \mathcal{O}(n)$$

CLT and asymptotic consequences

Under Hypothesis A and a similar one replacing p by $\hat{p}_n$, then

$$\Lambda_n^{1/2}\left(\hat{p}_n - p\right) \quad \sim \quad \mathcal{N}\left(0, \mathbf{I_d}\right)$$

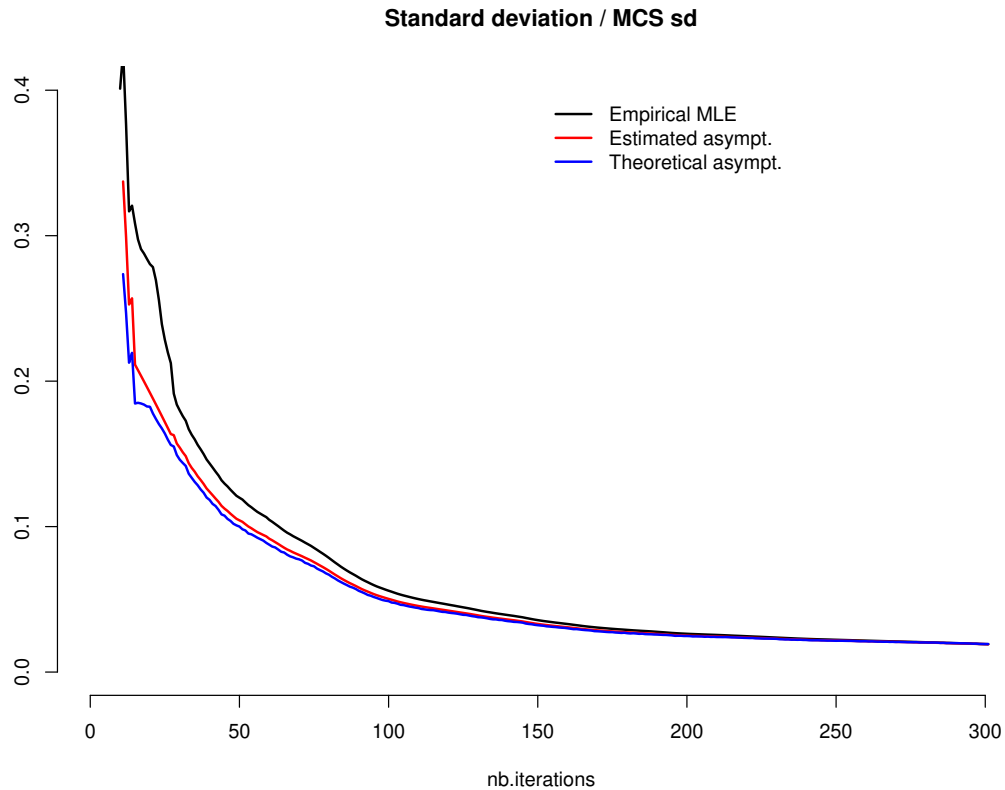
Variance reduction: $\Lambda_n^{-1} \leq \left(\rho_{n-1}^{med}\right)^2 \frac{p}{1-p} \mathrm{Var}^{MCS}$

Asymptotic robustness: $\lim_{p \to 0} \mathrm{CV}\left[\hat{p}_n\right] < \infty$ for $\to \infty$

Illustration

Asymptotic variance estimated using (rule of thumb)

- unbiased estimates of the weights
- MLE $\hat{p}_n$



Points to enlight / limitations

Current results on asymptotism seem difficult to ensure in practice

Looking for easier conditions based on the behavior of ρ_n: difficult to have in very generalized cases without information on G

When asymptotism begin?

We would prefer importance sampling approaches to get

- always unbiased estimators
- non-asymptotic variance reduction
- possibilities of re-estimation without new runs of G **(like the computation of bounds)**

Furthermore, at step i, we would prefer to sample $\boldsymbol{\nu}_i$ from an arbitrary importance pdf f_{i-1} with support $\mathbb{U}_{i-1}$

Defining Weighted Importance Sampling estimators (WISE)

Replacing the weights ($\tilde{\omega}_i(p)$) by deterministic weights (ω_i) (summing to n)

$$\hat{p}_n = \frac{1}{n}\sum_{i=1}^{n} \omega_i \bar{p}_i$$

with the local unbiased estimator

$$\bar{p}_i = p_{i-1}^{-} + \frac{\xi_{\nu_i}}{f_{i-1}(\boldsymbol{\nu}_i)}$$

To have $\bar{p}_i \in [p_{i-1}^{-}, p_{i-1}^{+}]$, one must ensure

$$f_{i-1}(\boldsymbol{\nu}) \geq \frac{1}{p_{i-1}^{+} - p_{i-1}^{-}} \quad \forall \nu \in \mathbb{U}_{i-1}$$

When f_i is uniform, $\bar{p}_i$ is the local estimator used in the expression of the MLE

Properties of WISE (1)

Variance

- Denote $c_0 = 0$ and $c_n = \mathrm{E}_{\mathcal{F}_n}\left[\frac{p_n^-}{p} + \frac{1-p_n^+}{1-p} - \frac{p_n^-(1-p_n^+)}{p(1-p)}\right] \quad \forall n > 1$
- Denote $\Lambda_n(f_n) = \mathrm{E}_{\mathcal{F}_n}\left[(p_n^+ - p_n^-)\left[\int_{\mathbb{U}_n} \frac{\mathbb{1}_{\{\tilde{G}(\nu)\leq 0\}}}{f(\nu)}\, d\nu - (p - p_n^-)\right]\right]$

Then

$$\mathrm{Var}\left[\hat{p}_n\right] = \mathrm{Var}^{unif}\left[\hat{p}_n\right] + \frac{1}{n}\sum_{i=1}^{n}\omega_i^2\Lambda_{i-1}(f_{i-1})$$

where

$$\mathrm{Var}^{unif}\left[\hat{p}_n\right] = \mathrm{Var}^{MCS}\frac{1}{n}\sum_{i=1}^{n}\omega_i^2\left(1 - c_{i-1}\right)$$

is the variance of $\hat{p}_n$ when f_{i-1} is chosen uniform

---

Example: if f_i is uniform and ω_i is constant, then

$$\mathrm{Var}\left[\hat{p}_n\right] = \frac{p(1-p)}{n}\left(1 - \frac{1}{n}\sum_{i=1}^{n}c_{i-1}\right)$$

Properties of WISE (2)

If at each iteration i f_{i-1} is chosen such that $\Lambda_{i-1}(f_{i-1}) \leq 0$, then

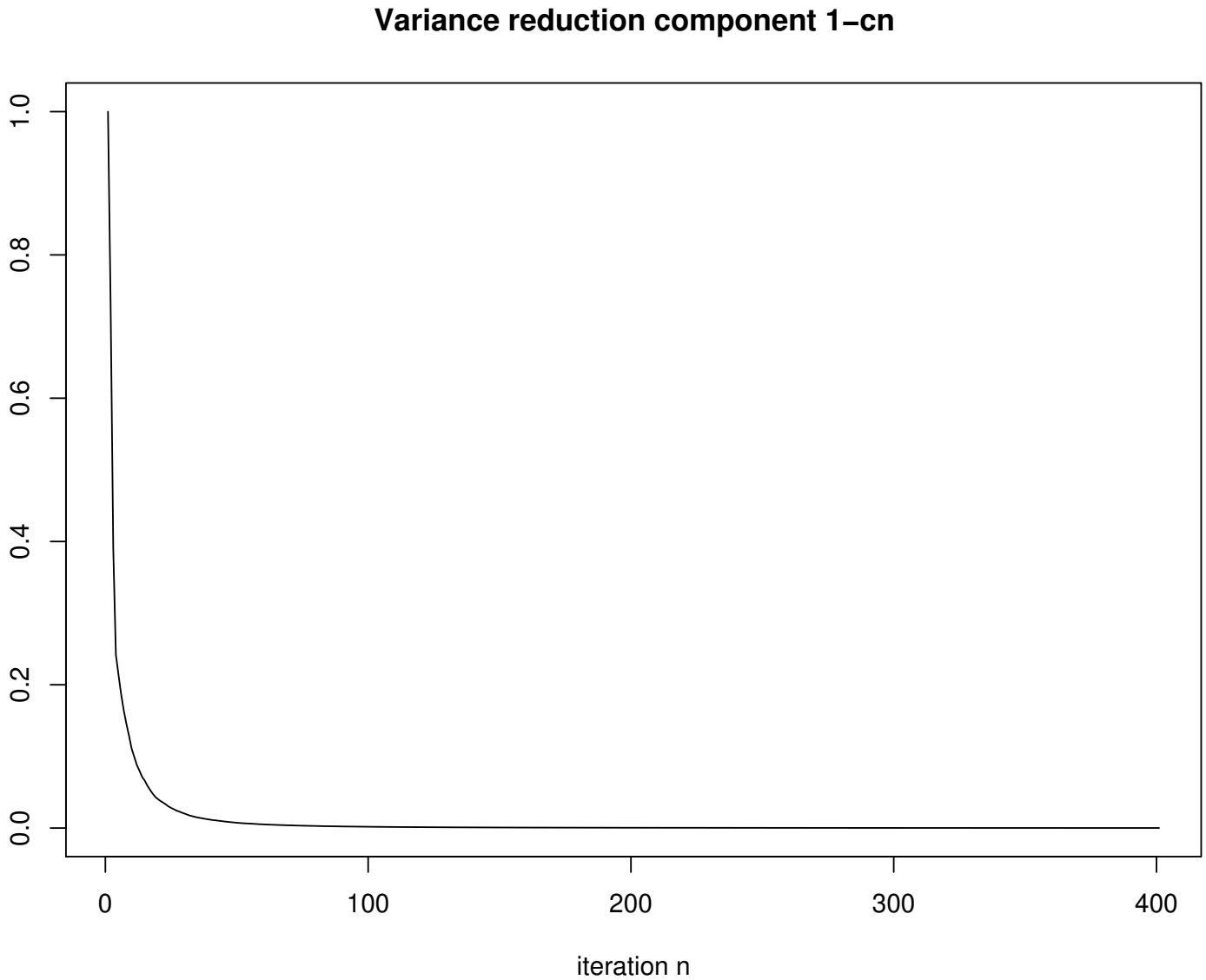
1. $\hat{p}_n$ is (non-asymptotically) **robust**:

$$\lim_{p \to 0} \mathrm{CV}\left[\hat{p}_n\right] \quad < \quad \infty$$

2. the variance of $\hat{p}_n$ is always smaller than the Monte Carlo variance

An unbiased, consistent estimator of $\mathrm{Var}\left[\hat{p}_n\right]$ can be found when f_i is uniform (no other results yet ; I'm working on it)

Illustration: $1 - c_n$ in function of n ($d = 2$)



Two linked difficulties

1. **Boundness:** one cannot theoretically ensure that WISE remains in $(p_{n-1}^{-}, p_{n-1}^{+})$
   - *Idea:* starting the estimation after n_ϵ^*, $0 < \epsilon \ll 1$ where

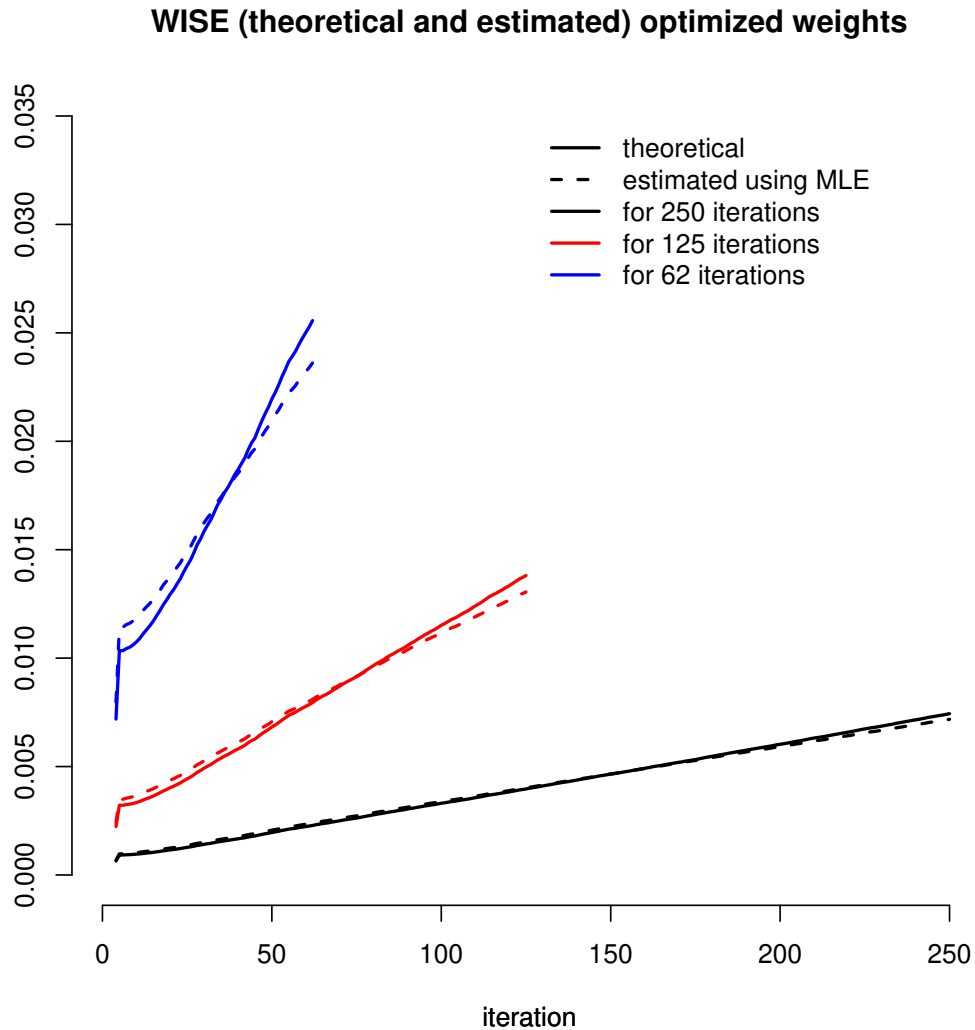
$$\left| \frac{\partial LCI_n(p_n^-)}{\partial n}(n_\epsilon^*) \right| \quad = \quad \epsilon \quad \text{(approximated using a finite difference approach)}$$

   - The main bias w.r.t. bounds due to first iterations should vanish
   - Can be used too in the MLE computation...
2. **Selecting the weights:** variance minimizers

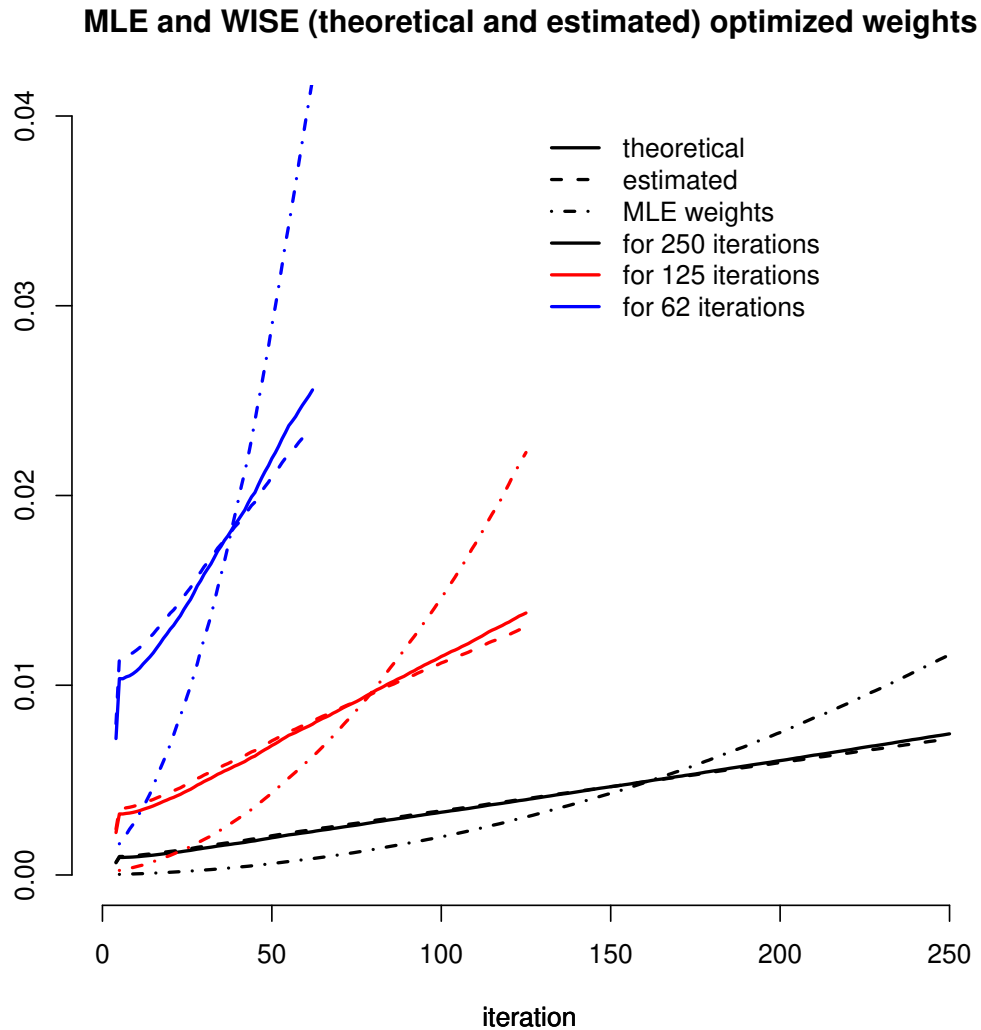
$$\omega_i^* \quad \propto \quad \left[(1 - c_{i-1}) \sum_{k=1}^{n} (1 - c_{k-1})^{-1} \right]^{-1} \qquad \text{(uniform case)}$$

   - Can be easily generalized to any importance density
   - Such weights increase too
   - Must be estimated

Example in dimension 2 ($p = 5.10^{-2}$)



Example in dimension 2 ($p = 5.10^{-2}$)



Some issues to consider

In addition to increasing our simulation studies (!)....

1. Eliciting good importance densities
2. Behavior of estimators (compound Poisson, asymptotic normality...?)
3. Starting / stopping the useful DOE for a good statistical estimation
4. Preserving monotony through transformations of inputs
5. Computational geometry issues

Eliciting importance densities: a first issue

At each step i, $f_{i-1}(\boldsymbol{\nu})$ should be elicited under the following constraints:

1. $\bar{p}_i \in \left[p_{i-1}^-, p_{i-1}^+\right]$ (local estimator)
2. $\hat{p}_i \in \left[p_{i-1}^-, p_{i-1}^+\right]$ (global estimator)
3. $\mathrm{Var}\left[\bar{p}_i | \boldsymbol{\nu}_1, \ldots, \boldsymbol{\nu}_{i-1}\right]$ is minimized

A good choice of f_i should be related too

- to a **prior idea** of the failure surface $S = \{\boldsymbol{\nu} \in \mathbb{U}_{i-1},\ \tilde{G}(\boldsymbol{\nu}) = 0\}$
  - using a sequential approximation of S (**kriging**, helped by evolving bounds... ?)
  - using some knowledge about the less probable regions where S is located

A naive approach when there is little idea about the failure surface

1. Sample N points $(\boldsymbol{\nu}_{i,1}, \ldots, \boldsymbol{\nu}_{i,N}) \in \mathbb{U}_{i-1}^{N}$
2. Estimate a nonparametric distribution of these points (clusters classification?)
3. Consider the modal values as good indicators of areas of highest concentration
   - favor "widest" areas between $\mathbb{U}_{i-1}^{-}$ and $\mathbb{U}_{i-1}^{+}$
   - one may spread lots of points inside!
4. Elicit f_{i-1} such that especially favoring points around modal values

Another close idea (recently inspired by A. Bar-Hen):

- Build a Voronoi tesselation of $\mathbb{U}_{i-1}$ using a small number of points (germs) spread from a maximin algorithm
- Consider the widest cell $C_{i-1}^{\max}$ as a good indicator of the widest possible space between $\mathbb{U}_{i-1}^{-}$ and $\mathbb{U}_{i-1}^{+}$
- Elicit

$$f_{i-1}(\boldsymbol{\nu}) \quad = \quad (1-\lambda)\mathbb{1}_{\{\boldsymbol{\nu} \in C_{i-1}^{\max}\}} + \lambda \mathbb{1}_{\{\boldsymbol{\nu} \in \mathbb{U}_{i-1}/C_{i-1}^{\max}\}}$$

with $0 < \lambda \ll 1$

Asymptotic behavior of WISE: a second issue

Progressive WISE: replacing the weights (ω_i) by increasing deterministic weights $(\hat{\omega}_i)$ such that one can write

$$\hat{p}_n \quad = \quad \hat{\omega}_n \hat{p}_{n-1} + (1 - \hat{\omega}_n)\bar{p}_n$$

Transformation

$$\omega_i \quad = \quad (1 - \hat{\omega}_{i-1}) \prod_{k=i}^{n} \hat{\omega}_k$$

Sequentialness is important to use martingale arguments $\Rightarrow$ proving CLT:

$$\prod_{i=1}^{n} \hat{\omega}_i^{-1} (\hat{p}_n - p) \quad \text{is a martingale}$$

Similar conditions than before (LLN on the c_i in the uniform case and their empirical estimates)

Starting / stopping statistical estimation : a third issue

Evolving bounds are practical too to get **concentration inequalities** (Azuma-Hoeffding) independent of the distribution of $\hat{p}_n$

Conservativeness: the probability $p_n(\epsilon)$ of underestimating $(1-\epsilon)p$ after n steps takes the general form (MLE/WISE)

$$p_n(\epsilon) \quad \leq \quad \exp\left(-2(n\epsilon)^2 \left[\sum_{k=1}^{n} \rho_k^2\right]^{-1}\right) \tag{1}$$

Of particular interest when the bounds are close (ie, they are not trivial)

Might be of help to select the first points to consider in a statistical estimation

General conditions for preserving (increasing) monotony from $\boldsymbol{D}$ to $\mathbb{U}$: a fourth issue

---

Obvious lemma on conditioning

Assume that for $k = 2, \ldots, d$, there exists a function 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 an increasing function of $X_1, \ldots, X_{k-1}$.

- any standard bi-elliptical input distribution with positive correlation coefficient preserves monotony

---

Condition on normal copulas with correlation matrix ρ [Chen 2009]

Denote $\boldsymbol{\rho}^k = (\boldsymbol{\rho}_{i,j})$, $1 \leq i, j \leq k \leq d$ and let $|\boldsymbol{\rho}_{i,j}^{(k)}|$ be the cofactor of $\boldsymbol{\rho}_{i,j}$ in $\boldsymbol{\rho}^k$;

(i) $\boldsymbol{\rho}$ must be a completely positive matrix,

(ii) one must have $|\boldsymbol{\rho}_{i,j}^{(k)}| \leq 0 \ \forall i > j$,

Probably true for elliptical copulas

Computational geometry problems

- Volume calculation (union of hyperrectangles)
  - Improving sweepline algorithms solving Klee's measure problem in due time in high dimension (difficult subject!)
- Best deterministic/stochastic strategies for removing much space in first iterations?
  - Using a Bayesian framework since many engineers may have an idea of an upper bound for p (possibly conditional to some input events)
  - Recent directional sampling heuristics (Limbourg et al. 2010) seem to work well

Connections with other statistical/probabilist issues

- Coupling with other methods (kriging ...) when choosing sequential importance functions
- Other definitions of $\hat{p}_n$ (e.g., splitting) ?

Tools from martingale theory (jonction point between statistics and probabilities)

Some words for a temporary conclusion....

Focusing only on general/naive results

Even though bounds are too conservative and does not move a lot after some iterations, they can help to get faster and more robust estimators than the ones currently used

Until now only tested in small dimensions (slowness of computational geometry algorithms)

Supplementary hypotheses should be made to explore faster the uniform space at the beginning of the computation

We suggest considering monotony to adapt sequential approaches

- using monotony as a prior knowledge

Main bibliography

- Bousquet, N., de Rocquigny, E. (2010). Bounding and estimating exceedance probabilities under high computational cost: taking advantage of monotony, *preprint under progress*.
- Genest, C., Favre, A.N. (2007). Everything you always wanted to know about copula modeling but were afraid to ask. *J. Hydr. Engin.*, 12: 347-368.
- Gille-Genes, A. (1999). Utilisation des méthodes numériques probabilisites dans les applications au domaine de la fiabilité des structures. Ph.D. Paris VI.
- Heijmans, R.D.H., Magnus, J.R. (1986). On the first-order efficiency and asymptotic normality of MLE obtained from dependent observations. *Statistica Neerlandica*, 40: 169-187.
- Kroese, D.P., Rubinstein, R.Y. (2007). *Simulation and the Monte Carlo Method (2nd ed.)*. Wiley.
- L'Ecuyer, P., Demers, V., Tuffin, B. (2007). Rare events, Splitting and Quasi-Monte Carlo. *ACM Transactions on Modeling and Computer Simulation (TOMACS)*, 17, nř9.
- Rajabalinejad, M., Van Gelder, P.H., Vrijling, J.K. (2008). Improved Dynamic Bounds in Monte Carlo Simulations. *AIAA 2008*.
- de Rocquigny, E. (2009). SStructural reliability under monotony: Properties of FORM, simulation or response surface methods and a new class of Monotonous Reliability Methods (MRM). *Structural Safety*, 31: 363-374.

Supplementary precisions

- Transforms from $\boldsymbol{D}$ to $\mathbb{U}$
- Sweepline algorithms

From $\boldsymbol{D}$ to $\mathbb{U}$: transforms (1)

(example 1) Stochastic ordering: Rosenblatt's transform

$$\Psi_{F_{\mathbf{X}}}(\mathbf{X}, \mathbf{U}) = \left(F_1(X_1, U_1), \ldots, F_{d|1,\ldots,d-1}(X_d, U_d)\right)$$

(example 2) Normal copula for $\mathbf{X}$: Nataf's transform

1. Let $\mathbf{Y} = \Phi^{-1}\left(F_{X_1}(X_1), \ldots, F_{X_d}(X_d)\right)$: gaussian vector/copula with correlation matrix $\boldsymbol{\rho}$
2. Then $\boldsymbol{\nu} = \Phi(\boldsymbol{\Gamma}\mathbf{Y})$ with $\boldsymbol{\Gamma}^{-t}\boldsymbol{\Gamma}^{-1} = \boldsymbol{\rho}$

Sweepline algorithms

Sweepline (or *plane sweep*) algorithms are commonly used to jointly detect and sort intersections between segments

A d-dimensional volume is calculated recursively by calculating all $n-1$-dimensional “slices" of the d-th dimension