Point process-based approaches for robust reliability analysis of systems modeled by expensive simulators

G. Perrin[1]

Journée Utilisateurs OpenTURNS #15 | June 9th 2022

[1]COSYS, Université Gustave Eiffel, 77420 Champs-sur-Marne, France

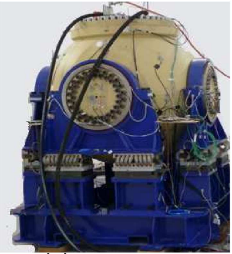
Introduction

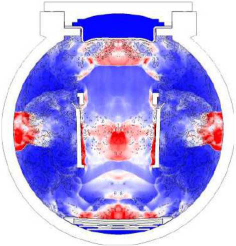
- Simulation plays a key role in the reliability analysis of complex systems.
- Most of the time, these analyses can be reduced to estimating the probability of occurrence of an undesirable event, using a stochastic model of the system.
- If the considered event is rare, sophisticated sample-based procedures are generally introduced to get a relevant estimate of the failure probability.

Problematic

Based on a reduced number of model evaluations (**costly simulators**), how to **bound** this failure probability with a prescribed confidence (**robust estimation**) ?

A first example


(a) Real tank


(b) Hydrodynamics

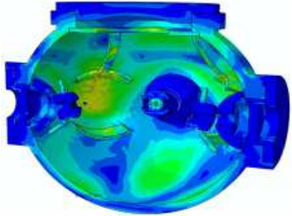

(c) Structure dynamics

FIGURE: Pressure tank under dynamic pressure

Problematic

How to certify that the maximum value in time and space of the cumulative equivalent plastic strain is less than a prescribed value ?

Outline of the presentation

Point process-based approaches for rare events
General framework

The reliability of a system (physical, chemical, mechanical, financial) refers to several quantities :

- $\boldsymbol{x} = (x_1, \ldots, x_d) \leftrightarrow$ parameterisation of the system, which is modelled as a **random vector**, whose PDF is noted $f_{\boldsymbol{X}}$,
- $y \leftrightarrow$ quantity of interest for the monitoring of the good functioning of the considered system,
- $S = 0 \leftrightarrow$ threshold not to be exceeded.

To guarantee the correct functioning of the system, we first need to **calculate** the **probability of failure**, noted p_f, and verifying :

$$p_f := \mathbb{P}(y(\boldsymbol{x}) < 0),$$

and then to **decide** whether this value is admissible or not (with regard to safety standards for example).

Point process-based approaches for rare events

Extreme value theory

Extreme value theory can be seen as a "parametric" post-processing of a very large number of evaluations of y at randomly and independently drawn points $\boldsymbol{x}$, for the evaluation of p_f. The methods which will be presented in the following differ from this by the fact that we assume here :

- that an evaluation of $y(\boldsymbol{x})$ is relatively "**expensive**" (financially for an experiment or numerically for a simulation),
- that in the initial state, **no evaluation** of y has been carried out.

Problematic

The objective is to define a **sequence** (in the sense that past results can be used to define new evaluation points) of n input points $\boldsymbol{x}$, and therefore of n evaluations of y, of **minimum size**, allowing the **best estimation** of p_f.

Point process-based approaches for rare events
Several thresholds for risk assessment

$$p_f := \mathbb{P}(y(\boldsymbol{x}) < 0),$$

- The system is "sufficiently" safe if $p_f \leq \alpha$.
- If $\widehat{p}_n$ is a statistical **estimator** of p_f based on n evaluations of y, the system can be considered as "sufficiently" safe if $\widehat{p}_n + c(n,\alpha,\beta) \leq \alpha$, where $c(n,\alpha,\beta)$ is adjusted to avoid false certification with high probability $1-\beta$:

$$\max_{p_f \geq \alpha} \mathbb{P}(\widehat{p}_n + c(n,\alpha,\beta) \leq \alpha) \leq \beta.$$

The safety assessment relies on several constants

→ $S = 0 \leftrightarrow$ threshold on y not to be exceeded,

→ $\alpha \leftrightarrow$ "acceptable" risk (in reference to S),

→ $\beta \leftrightarrow$ confidence level (replacement of p_f by $\widehat{p}_n$),

→ $c(n,\alpha,\beta) \leftrightarrow$ security margin (in reference to $\widehat{p}_n, S, \alpha, \beta$).

Point process-based approaches for rare events
The Monte Carlo case

- Let $\boldsymbol{x}_1, \dots, \boldsymbol{x}_n$ be n i.i.d. realizations of $\boldsymbol{x}$,
- $\widehat{p}_n := \frac{1}{n} \sum_{i=1}^{n} 1_{y(\boldsymbol{x}_i)<0}$ is the classical Monte Carlo estimator of p_f.
- Noticing that $\sqrt{n}(\widehat{p}_n - p_f) \xrightarrow{\mathcal{L}} \mathcal{N}(0, p_f(1-p_f))$, it comes that

$$c(n, \alpha, \beta) = \phi_{1-\beta}\sqrt{\alpha(1-\alpha)/n}$$

is a natural choice to get (asymptotically) the desired certification criterion, with $\phi_{1-\beta}$ the $(1-\beta)$ quantile of a standard Gaussian r.v :

$$\max_{p_f \geq \alpha} \mathbb{P}(\widehat{p}_n + c(n, \alpha, \beta) \leq \alpha) \underset{n \to +\infty}{\longrightarrow} \beta.$$

Point process-based approaches for rare events
The Monte Carlo case

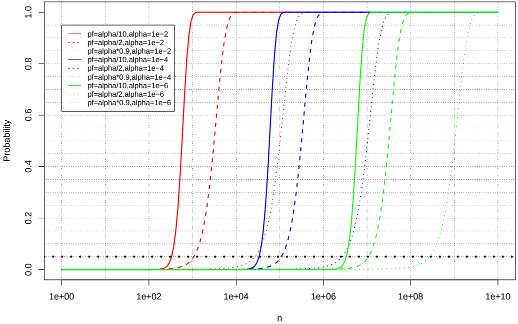


FIGURE: $n \mapsto \mathbb{P}(\widehat{p}_n + c(n, \alpha, \beta = 5\%) \leq \alpha)$

$\Rightarrow$ Many code evaluations are required to get a satisfying confidence on the results ($n \approx 10\alpha$ when $p_f \approx 0.1\alpha$, $n \approx 50\alpha$ when $p_f \approx 0.5\alpha$, $n \approx 10^3\alpha$ when $p_f \approx 0.9\alpha$).

Point process-based approaches for rare events
Point-process based estimator

As an alternative, let us consider the following algorithm :

- Let $\boldsymbol{x}_1, \dots, \boldsymbol{x}_Q$ be Q iid copies of $\boldsymbol{x}$ and $\mathcal{S} = \{y(\boldsymbol{x}_1), \dots, y(\boldsymbol{x}_Q)\}$.
- While $\max(y(\boldsymbol{x}_1), \dots, y(\boldsymbol{x}_Q)) \geq 0$:
  - → Find $q^{\mathsf{max}} = \arg\max_{1 \leq q \leq Q} y(\boldsymbol{x}_q)$,
  - → Redefine $\boldsymbol{x}_{q^{\mathsf{max}}}$ by $\boldsymbol{x} | y(\boldsymbol{x}) < y(\boldsymbol{x}_{q^{\mathsf{max}}})$,
  - → Add $y(\boldsymbol{x}_{q^{\mathsf{max}}})$ to $\mathcal{S}$.
- Return $s_1 \geq s_2 \geq \dots \geq s_M$ the M values of $\mathcal{S}$ that are greater than 0 so that :

$$p_f = \mathbb{P}(y(\boldsymbol{x}) < 0 | y(\boldsymbol{x}) < s_M) \times \mathbb{P}(y(\boldsymbol{x}) < s_M | y(\boldsymbol{x}) < s_M) \times \cdots \times \mathbb{P}(y(\boldsymbol{x}) < s_2 | y(\boldsymbol{x}) < s_1) \times \mathbb{P}(y(\boldsymbol{x}) < s_1),$$

(By construction, all these probabilities are likely to be close to $1 - \frac{1}{Q}$).

M and $N := \sharp\mathcal{S}$ are two **random quantities**, and we can define :

$$\widetilde{p}_N := \left(1 - \frac{1}{Q}\right)^M.$$

Point process-based approaches for rare events

Point-process based estimator

$$\widetilde{p}_N := \left(1 - \frac{1}{Q}\right)^M, \quad N := \sharp\, \mathcal{S}.$$

Properties

- $\mathbb{E}\left[\widetilde{p}_N\right] = p_f, \quad \mathrm{Var}(\widetilde{p}_N) = p_f^2(p_f^{-1/Q} - 1),$
- $\mathbb{E}\left[N\right] = Q(1 - \log(p_f)),$
- $\log\left(\widetilde{p}_N\right) \xrightarrow{\mathcal{L}} \mathcal{N}\left(\begin{array}{c} -Q\log(p_f)\log(1-1/Q), \\ -Q\log(p_f)\log(1-1/Q)^2 \end{array}\right) \approx \mathcal{N}\left(\log(p_f), -\log(p_f)/Q\right).$

We deduce that if $c(Q,\alpha,\beta) = \alpha\left(1 - \exp\left(\phi_\beta\sqrt{-\frac{\log(\alpha)}{Q}}\right)\right)$, then

$$\max_{p_f \geq \alpha} \mathbb{P}(\widetilde{p}_N + c(Q,\alpha,\beta) \leq \alpha) \underset{Q \text{ suff. high}}{\approx} \beta.$$

Point process-based approaches for rare events

The Monte Carlo case

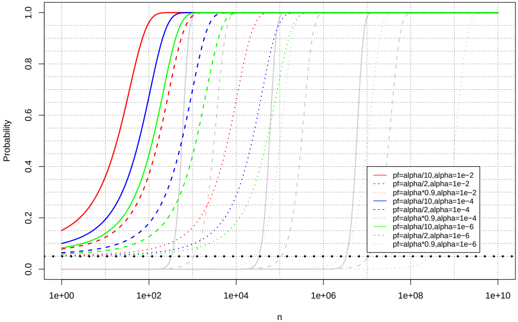


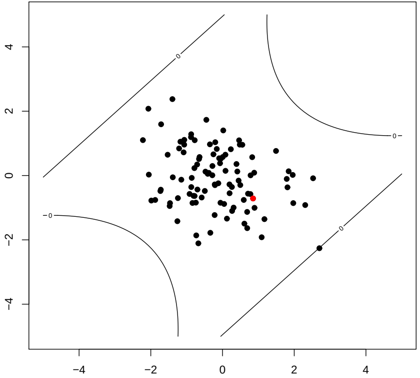
FIGURE: $n \mapsto \mathbb{P}(\widetilde{p}_N + c(Q, \alpha, \beta = 5\%) \leq \alpha)$, $n = Q(1 - \log(p_f))$.

$\Rightarrow$ Much less code evaluations are required to get the same confidence on the results, especially for low values of p_f.

Point process-based approaches for rare events

A simple example - $p_f \approx 0.0022$.

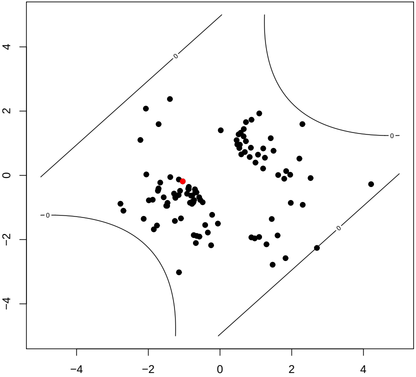
$Q = 100$. Red : current maximum at initial step (no move).



Point process-based approaches for rare events

A simple example - $p_f \approx 0.0022$.

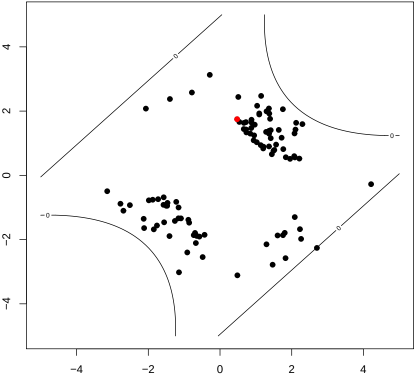
$Q = 100$. Red : current maximum after 100 steps.



Point process-based approaches for rare events

A simple example - $p_f \approx 0.0022$.

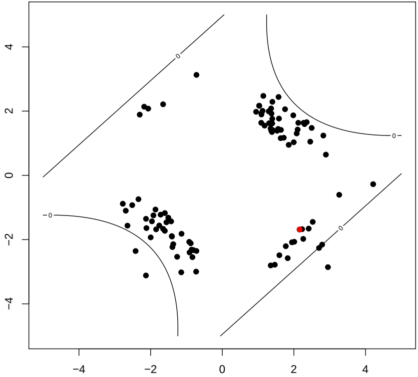
$Q = 100$. Red : current maximum after 200 steps.



Point process-based approaches for rare events

A simple example - $p_f \approx 0.0022$.

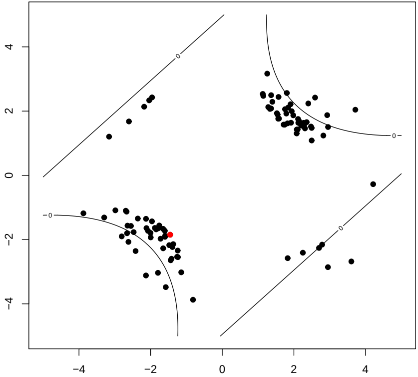
$Q = 100$. Red : current maximum after 300 steps.



Point process-based approaches for rare events

A simple example - $p_f \approx 0.0022$.

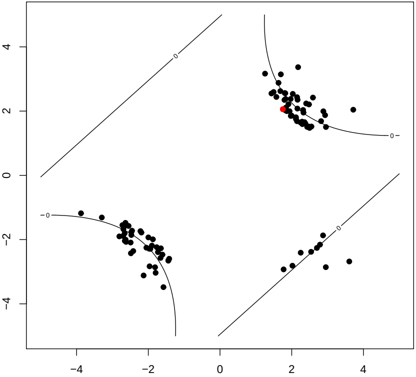
$Q = 100$. Red : current maximum after 400 steps.



Point process-based approaches for rare events

A simple example - $p_f \approx 0.0022$.

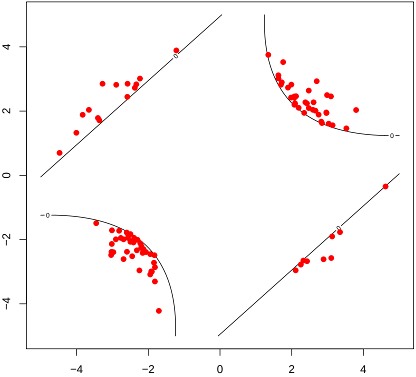
$Q = 100$. Red : current maximum after 500 steps.



Point process-based approaches for rare events

A simple example - $p_f \approx 0.0022$.

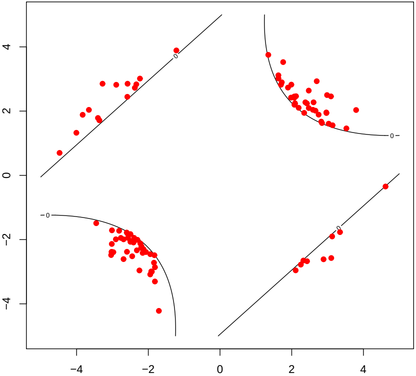
$Q = 100$. After $M = 601$ steps, all the points are below 0.



Point process-based approaches for rare events

A simple example - $p_f \approx 0.0022$.

$Q = 100$. After $M = 601$ steps, all the points are below 0.



$$\widetilde{p}_N \approx \left(1 - \frac{1}{Q}\right)^M \approx 0.0024.$$

Point process-based approaches for rare events
Summary

Monte Carlo case

- $\widehat{p}_n := \frac{1}{n}\sum_{i=1}^{n} 1_{y(\boldsymbol{x}_i)<0}$, $\boldsymbol{x}_1, \dots, \boldsymbol{x}_n$ iid copies of $\boldsymbol{x}$, $\widehat{\delta}_n^2 \approx \frac{p_f^{-1}}{n}$
- Around p_f^{-1} code evaluations to get one realization of $\boldsymbol{x}$ in the failure domain.

(+) Easy to implement. (-) Costly.

Moving particle domain

- $\widetilde{p}_N := \left(1 - \frac{1}{Q}\right)^M$, $(y(\boldsymbol{x}_i))_{i=1}^N$ is a decreasing random walk (strong link with Poisson processes), $\widetilde{\delta}_n^2 \approx \frac{-\log(p_f)}{Q}$
- Around $Q(1 - \log(p_f))$ code eval. to get one real. of $\boldsymbol{x}$ in the failure domain.

(+) Less costly. (-) Need for strategies to implement the random walk.

$\Rightarrow$ $N \approx Q(1 - T\log(p_f))$ in practice, with T a burn-in parameter (MCMC).

$\Rightarrow$ More details on practical implementation in the **Mistral** R packages.

Outline of the presentation

Robustness and sensitivity analyses

Context

- The value of p_f completely depends on the PDF of $\boldsymbol{x}$.
- Given a risk α and a confidence level β, we proposed two estimators to numerically assess the system reliability. Two cases may occur :
  1. p_f seems to be higher than α (or at least there are too much uncertainty on the fact that $p_f < \alpha$) (**"negative"** configuration)

  $\Rightarrow$ what are the model inputs whose variability has to be reduced in priority to decrease p_f (**sensitivity**) ?
  2. p_f seems to be smaller than α with a reasonable confidence (**"positive"** configuration)

  $\Rightarrow$ what are the model inputs whose distribution has to be particularly well-characterized for the available estimate to be realistic (**robustness**) ?

Each code evaluation being costly, how could we **post-process** the samples generated during the estimation of p_f to answer to these two questions (**no additional cost**) ?

Robustness and sensitivity analyses
Sensitivity

- To find the quantities having the **most influence** on p_f, it is interesting to quantify the impact on p_f of fixing x_i at a certain value $x_i^\star$.
- The larger $(p_f - \mathbb{P}(y(\boldsymbol{x}) < 0 | \; x_i = x_i^\star))^2$ is, the more likely x_i is to play a role on p_f.
- By **averaging** this quantity over x_i, the quantity

$$\mathbb{E}\left[(\mathbb{P}(y(\boldsymbol{x}) < 0) - \mathbb{P}(y(\boldsymbol{x}) < 0 | x_i))^2\right] = \mathbb{V}_{x_i}\left[\mathbb{E}_{\boldsymbol{x}_{-i}}\left[1_{y(\boldsymbol{x})<0} | x_i\right]\right]$$

  allows us to analyse the **sensitivity** of p_f to each input of the model.
- Here we find the first order Sobol indices (and by extension total indices) of the function $1_{y(\boldsymbol{x})<0}$:

$$s_i := \frac{\mathbb{V}_{x_i}\left[\mathbb{E}_{\boldsymbol{x}_{-i}}\left[1_{y(\boldsymbol{x})<0} \mid x_i\right]\right]}{\mathbb{V}_{\boldsymbol{x}}\left[1_{y(\boldsymbol{x})<0}\right]}, \quad t_i := 1 - \frac{\mathbb{V}_{\boldsymbol{x}_{-i}}\left[\mathbb{E}_{x_i}\left[1_{y(\boldsymbol{x})<0} \mid \boldsymbol{x}_{-i}\right]\right]}{\mathbb{V}_{\boldsymbol{x}}\left[1_{y(\boldsymbol{x})<0}\right]}.$$

Robustness and sensitivity analyses
Sensitivity

To efficiently assess the values of s_i and t_i, it is interesting to notice that :

$$s_i = \frac{p_f}{1-p_f}\mathbb{V}_{x_i}\left[\frac{f_{x_i|y(\boldsymbol{x})<0}(x_i)}{f_{x_i}(x_i)}\right], \quad t_i = 1 - \frac{p_f}{1-p_f}\mathbb{V}_{\boldsymbol{x}_{-i}}\left[\frac{f_{\boldsymbol{x}_{-i}|y(\boldsymbol{x})<0}(\boldsymbol{x}_{-i})}{f_{\boldsymbol{x}_{-i}}(\boldsymbol{x}_{-i})}\right].$$

Hence, by estimating $f_{x_i|y(\boldsymbol{x})<0}$ and $f_{\boldsymbol{x}_{-i}|y(\boldsymbol{x})<0}$, using **kernel smoothing techniques** for instance ("simple post-processing of the failure points"), it is possible to quantify and compare the influence on p_f of each model input.

Remark

When p_f is small, the interest of such indices may be limited as the values of s_i are likely to be all close to 0, when the values of t_i are all likely to be close to 1 (a failure event is generally associated with a pathological combination of all the model inputs).

Robustness and sensitivity analyses
Robustness

Everything that has been presented so far is based on a **fixed** definition of the sources of uncertainty.

Let us now note that for any positive function $f_1, \dots, f_d$ defined on $\mathbb{R}$ and of integral 1, the quantity $h(f_1, \dots, f_d)$ such that :

$$h(f_1, \dots, f_d) := \int_{\mathbb{X}} 1_{y(\boldsymbol{x})<0} \prod_{j=1}^{d} f_j(x_j) d\boldsymbol{x}$$

defines the probability that $y(\boldsymbol{x})$ is smaller than 0, **under the condition** that the PDF of $\boldsymbol{x}$ is equal to $\prod_{j=1}^{d} f_j$.

Robustness and sensitivity analyses
Robustness

For $0 \leq \delta < 1$, we then note $\mathcal{F}_i(\delta)$ a set of PDFs corresponding to **perturbations** of "amplitude δ" of f_{x_i} (e.g., modification of the mean and variance for a Gaussian variable), and :

$$p^\delta := \max_{f_j \in \mathcal{F}_j(\delta),\ 1 \leq j \leq d} h(f_1, \dots, f_d),$$

$$p_i^\delta := \max_{f_i \in \mathcal{F}_i(\delta)} h(f_{x_1}, \dots, f_{x_{i-1}}, f_i, f_{x_{i+1}}, \dots, f_{x_d}),$$

$$p_{-i}^\delta := \max_{f_j \in \mathcal{F}_j(\delta),\ 1 \leq j \leq d,\ j \neq i} h(f_1, \dots, f_{i-1}, f_{x_i}, f_{i+1}, \dots, f_d).$$

By construction :

- → p^δ corresponds to the **worst case** by perturbing all PDFs,
- → p_i^δ by perturbing only that of x_i,
- → p_{-i}^δ by perturbing all PDFs of the components of $\boldsymbol{x}$ except the i^{th}.

Robustness and sensitivity analyses
Robustness

The influence of each input on the difference $p^\delta - p_f$ can thus be characterised by comparing the following two indices :

$$\varsigma_i^\delta := \frac{p_i^\delta - p_f}{p^\delta - p_f}, \quad \tau_i^\delta := 1 - \frac{p_{-i}^\delta - p_f}{p^\delta - P_f}.$$

- ς_i^δ characterises the percentage increase due to **individual** effects,
- τ_i^δ characterises the percentage increase due to **individual and coupled** effects.

The value of δ can be chosen to guarantee a given value of $p^\delta - p_f$ (for example $p^\delta = 2 \times p_f$ in the following - connections with the infogap theory).

It is possible to estimate all of these quantities **without new calls** to the code by using importance sampling-based methods (the variance may however be relatively high).

Robustness and sensitivity analyses

A simple example

$$y(\boldsymbol{x}) = 250 - (1 + x_1)(5 + x_2)(10 + x_3), \quad x_i \sim \mathcal{N}(0,1),$$

and we are interested in assessing $p_f = \mathbb{P}(y(\boldsymbol{x}) < 0)$. Using the former Moving Particle approach, we find : $\mathbb{P}(\mathbb{P}(y(\boldsymbol{x}) < 0) \in [0.0080; 0.0092]) \approx 95\%$.

		$\widehat{s}_i(\%)$	$\widehat{t}_i(\%)$	$\widehat{\varsigma}_i^\delta(\%)$	$\widehat{\tau}_i^\delta(\%)$
i=1	Reference	[1 ;21]	[94 ;107]	[57 ;61]	[70 ;73]
	Code+MP	[6 ;18]	[97 ;99]	[54 ;64]	[67 ;75]
i=2	Reference	[-14 ;10]	[74 ;87]	[18 ;21]	[28 ;31]
	Code+MP	[0 ;1]	[41 ;81]	[16 ;22]	[25 ;32]
i=3	Reference	[-9 ;8]	[47 ;56]	[7 ;9]	[12 ;15]
	Code+MP	[0 ;0]	[41 ;78]	[6 ;11]	[10 ;18]

- $\delta = 0.05$ so that $p^\delta \approx 2p_f$. The whole procedure is repeated 100 times.
- "Reference" : MC based on 6.10^6 evaluations.
- "Code+MP" relies on approximately 7200 code evaluations.

Outline of the presentation

Coupling GPR and point process-based approaches
Context

$$p_f := \mathbb{P}(y(\boldsymbol{x}) < 0).$$

- When confronted to **very expensive deterministic "black box" codes**, the former point-process based approach may also be too costly.

⇒ In that case, the calculation of p_f generally relies on the replacement of y by a **surrogate model**.

- We focus here on the Gaussian process regression (GPR), which models y as a particular realization of a Gaussian process $Y \sim \text{GP}(\mu, \Sigma)$.
- Under that formalism, $p_f = \mathbb{P}\left(Y(\boldsymbol{x}) < 0 \mid Y = y\right)$.

⇒ p_f is **a particular realization** of the random variable :

$$P_f^Y := \mathbb{P}\left(Y(\boldsymbol{x}) < 0 \mid Y\right).$$

Coupling GPR and point process-based approaches

Surrogate modeling and reliability analysis

- To correctly anticipate the risks of deterioration of the system, we therefore need to work on the construction of **confidence bounds** to failure probability estimates.
- Instead of working on the estimation of the mean value of P_f^Y, which is *a priori* as likely to overestimate as to underestimate p_f, we would like to construct a robust estimator $\widehat{Q}_{\alpha,\beta}$ of the (1-α) quantile q_α of P_f^Y, so that :

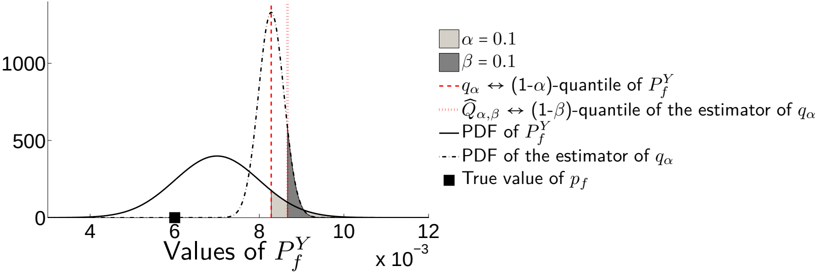
$$\mathbb{P}_Y(P_f^Y < q_\alpha) = 1 - \alpha,$$

$$\mathbb{P}_{\widehat{Q}_{\alpha,\beta}}\left(\mathbb{P}_Y\left(P_f^Y \leq \widehat{Q}_{\alpha,\beta} \mid \widehat{Q}_{\alpha,\beta}\right) \geq 1 - \alpha\right) \geq 1 - \beta.$$

Coupling GPR and point process-based approaches

Surrogate modeling and reliability analysis

$$\mathbb{P}_{\widehat{Q}_{\alpha,\beta}}\left(\mathbb{P}_Y\left(P_f^Y \leq \widehat{Q}_{\alpha,\beta} \mid \widehat{Q}_{\alpha,\beta}\right) \geq 1-\alpha\right) \geq 1-\beta.$$



- α characterizes the risk associated to the replacement of y by Y,
- β controls the fact that only finite-dimensional samples of $Y(\boldsymbol{x})$ are available for its construction.

Coupling GPR and point process-based approaches

New objectives

For $\alpha, \beta \in (0, 1)$ and a fixed number of evaluations of y,

- **First objective** : propose an algorithm allowing us to construct this estimator. Key elements :
  1. order statistics (Wilks' quantile type),
  2. the Gaussian process regression formalism,
  3. the former point-process sampling method.

Coupling GPR and point process-based approaches

New objectives

For $\alpha, \beta \in (0,1)$ and a fixed number of evaluations of y,

- **First objective** : propose an algorithm allowing us to construct this estimator. Key elements :
  1. order statistics (Wilks' quantile type),
  2. the Gaussian process regression formalism,
  3. the former point-process sampling method.
- **Second objective** : propose a strategy adapted to the former algorithm to sequentially minimize the dependence of $\widehat{Q}_{\alpha,\beta}$ on the replacement of y by Y, while managing the cases where :
  1. no point of the initial experimental design for the construction of Y belongs to the failure domain ,
  2. the failure domain is multimodal.

Coupling GPR and point process-based approaches

New objectives

For $\alpha, \beta \in (0,1)$ and a fixed number of evaluations of y,

- **First objective** : propose an algorithm allowing us to construct this estimator. Key elements :
  1. order statistics (Wilks' quantile type),
  2. the Gaussian process regression formalism,
  3. the former point-process sampling method.
- **Second objective** : propose a strategy adapted to the former algorithm to sequentially minimize the dependence of $\widehat{Q}_{\alpha,\beta}$ on the replacement of y by Y, while managing the cases where :
  1. no point of the initial experimental design for the construction of Y belongs to the failure domain ,
  2. the failure domain is multimodal.

Due to time constraints, only the first objective will be detailed in this presentation.

Coupling GPR and point process-based approaches

Initial exploration of the input space

Context reminder

- Input random vector : $\boldsymbol{x} \in \mathbb{X} \subset \mathbb{R}^d$ with PDF $f_{\boldsymbol{x}}$,
- Quantity of interest : $\boldsymbol{x} \mapsto y(\boldsymbol{x}) \in \mathbb{R}$,
- Failure probability : $p_f = \mathbb{P}(y(\boldsymbol{x}) < 0)$.

Gaussian process regression

- Model y has been evaluated in ℓ (the value of ℓ is assumed relatively small) points of $\mathbb{X}$, $\boldsymbol{x}^{(1)}, \dots, \boldsymbol{x}^{(\ell)}$ (space filling LHS).
- y is seen as a sample path of a Gaussian process defined on $(\Omega, \mathcal{A}, \mathbb{P})$.
- Let $Y \sim \mathrm{GP}(\mu, \Sigma)$ be this Gaussian process conditioned by the ℓ available code evaluations.

Coupling GPR and point process-based approaches
Order statistics (1/2)

- $Y_1, \dots, Y_m$ are $m \geq 1$ independent copies of Y,
- $\mathcal{X}_1^n, \dots, \mathcal{X}_m^n$ are $m \geq 1$ independent copies of a random set $\mathcal{X}^n$ of $n > 1$ points chosen (independently or not) in $\mathbb{X}$,
- $\widehat{P}_j := \widehat{P}_f^{Y_j, \mathcal{X}_j^n}$ is an **estimator** of p_f relying on the projection of Y_j in the n points of $\mathcal{X}_j^n$.

These estimators $\widehat{P}_1, \dots, \widehat{P}_m$ are supposed to be sorted in **ascending order**. From basic statistics, for $1 \leq j \leq m$ and $\alpha \in (0,1)$, we therefore have :

$$\mathbb{P}(\widehat{P}_j > q_\alpha) = \sum_{u=0}^{j-1} \binom{m}{u} (1-\gamma)^{m-u} \gamma^u, \quad \gamma := \mathbb{P}(\widehat{P}_f^{Y, \mathcal{X}^n} \leq q_\alpha).$$

Coupling GPR and point process-based approaches

Order statistics (2/2)

Noticing that $\gamma = \mathbb{P}(\widehat{P}_f^{Y,\mathcal{X}^n} \leq q_\alpha) \leq 1 - \alpha(1 - \mathbb{P}(\widehat{P}_f^{Y,\mathcal{X}^n} \leq P_f^Y \mid P_f^Y \geq q_\alpha)) =: \gamma_\star$, if we denote by $j^\star(\alpha,\beta)$ the minimal index such that

$$\sum_{u=0}^{j^\star(\alpha,\beta)-1} \binom{m}{u} (1-\gamma_\star)^{m-u} \gamma_\star^u \geq 1-\beta,$$

we obtain the two following results :

$$\mathbb{P}(\widehat{P}_{j^\star(\alpha,\beta)} > q_\alpha) \geq 1-\beta,$$

$$\mathbb{P}_{\widehat{P}_{j^\star(\alpha,\beta)}}\left(\mathbb{P}_Y\left(P_f^Y \leq \widehat{P}_{j^\star(\alpha,\beta)} \mid \widehat{P}_{j^\star(\alpha,\beta)}\right) \geq 1-\alpha\right) \geq 1-\beta.$$

which lead to the searched result when replacing $\widehat{P}_{j^\star(\alpha,\beta)}$ by $\widehat{Q}_{\alpha,\beta}$.

Coupling GPR and point process-based approaches
Choice of the estimator (1/2)

- As q_α is **unknown**, $\gamma_\star$ is **unknown** in the general case.
- Depending on the choice for the estimator of P_f^Y, asymptotic values can be proposed for $\gamma_\star$.
- For ex., if $Y(\omega)$ is a realization of Y and $\widehat{P}_f^{Y,\mathcal{X}^n}(\omega) = \sum_{i=1}^n 1_{Y(\boldsymbol{X}^{(i)};\omega)<0}/n$:

$$\sqrt{n}(\widehat{P}_f^{Y,\mathcal{X}_j^n}(\omega) - P_f^Y(\omega)) \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, P_f^Y(\omega)(1 - P_f^Y(\omega))\right) \quad \text{(CLT)}.$$

$\Rightarrow$ $\mathbb{P}(\widehat{P}_f^{Y,\mathcal{X}^n} \leq P_f^Y \mid P_f^Y \geq q_\alpha)$ tends to $1/2$ when n increases, which makes $\gamma_\star = 1 - \alpha(1 - \mathbb{P}(\widehat{P}_f^{Y,\mathcal{X}^n} \leq P_f^Y \mid P_f^Y \geq q_\alpha))$ tend to $1 - \alpha/2$.

However, when p_f is very small, to numerically calculate $\widehat{P}_f^{Y,\mathcal{X}^n}(\omega)$, we need to project Y in a very high number of points ($\approx 100/\widehat{P}_f^{Y,\mathcal{X}^n}(\omega)$), which is often not possible due to computational reasons (memory and conditioning problems).
$\Rightarrow$ **we need another estimator relying on a "reduced" number of code evaluations !**

Coupling GPR and point process-based approaches

Choice of the estimator (2/2)

- If $M(\omega)$ is a Poisson r.v. with parameter $P(-Q\log(\mathbb{P}_{\boldsymbol{x}}(Y(\boldsymbol{x};\omega)<0)))$,

$$\widehat{P}_f^{Y,\mathcal{X}_n}(\omega) := \left(1-\frac{1}{Q}\right)^{M(\omega)}$$

  defines an unbiased estimator of $P_f^Y(\omega)=\mathbb{P}_{\boldsymbol{x}}(Y(\boldsymbol{x};\omega)<0)$ such that :
  - $\gamma^\star$ becomes close to $1-\alpha/2$ when Q is high enough (asymptotic Gaussian behavior),
  - $Y(\omega)$ only needs to be projected in $\mathbb{E}[M(\omega)]=-Q\log(P_f^Y(\omega))$ points in average ($\ll 100/P_f^Y(\omega)$ for the former MC approach).
- Several realizations of $\widehat{P}_f^{Y,\mathcal{X}_n}$ can be obtained by launching **in parallel** the formerly presented Moving particle algorithm on several realizations of Y.

Coupling GPR and point process-based approaches
Practical implementation

Initialization

- Construct the GPR-based surrogate model associated with y based on ℓ evaluations of y, noted $Y \sim \mathrm{GP}(\mu, \Sigma)$.
- Choose risk level α and confidence level β (for instance $\alpha = 0.1$ and $\beta = 0.1$).
- Choose the initial sample size Q (for instance $Q = 100$).
- Choose the number of independent repetitions m (for $\alpha = \beta = 0.1$, $m \geq 45$).
- For $1 \leq j \leq m$ (this can be done in parallel) :
  - Sample Q independent realizations of $\boldsymbol{x}$, noted $\boldsymbol{x}(\omega_1), \ldots, \boldsymbol{x}(\omega_Q)$
  - Sample one realization of the Gaussian vector $(Y(\boldsymbol{x}(\omega_1)), \ldots, Y(\boldsymbol{x}(\omega_Q)))$, noted $(y_1, \ldots, y_Q)$
  - Define $Y_j(\omega) := Y \mid Y(\boldsymbol{x}(\omega_k)) = y_k, \; 1 \leq k \leq Q$
  - Set $n_{\text{iter}} = 0$, $\widehat{\mathcal{X}}^j = \{\boldsymbol{x}(\omega_1), \ldots, \boldsymbol{x}(\omega_Q)\}$, $\widehat{\mathcal{Y}}^j = \{y_1, \ldots, y_Q\}$.

Coupling GPR and point process-based approaches
Practical implementation

Iteration

For $1 \leq j \leq m$ (this can be done fully in parallel) :

- Set $z = \max(y_1, \ldots, y_Q)$, M^j = number of positive values of y
- While $z > 0$:
  - increment $n_{\text{iter}} = n_{\text{iter}} + 1$
  - draw at random a realization of $\boldsymbol{x}$, denoted by $\boldsymbol{x}^\star$
  - draw at random a realization of $Y_j(\boldsymbol{x}^\star)$, denoted by $y^\star$
  - If $y^\star < z$, actualize : $z = y^\star$, $M^j = M^j + 1$ if $y^\star > 0$, $Y_j(\omega) = Y_j(\omega) \mid Y_j(\boldsymbol{x}^\star) = y^\star$, $\widehat{\mathcal{X}}^j = \widehat{\mathcal{X}}^j \cup \{\boldsymbol{x}^\star\}$, $\widehat{\mathcal{Y}}^j = \widehat{\mathcal{Y}}^j \cup \{y^\star\}$.
- Compute $\widetilde{p}_j := \left(1 - \frac{1}{Q}\right)^{M^j}$.

Coupling GPR and point process-based approaches
Practical implementation

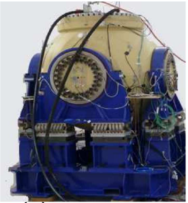
Iteration

For $1 \leq j \leq m$ (this can be done fully in parallel) :

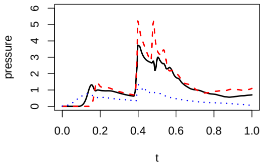
- Set $z = \max(y_1, \ldots, y_Q)$, M^j = number of positive values of y
- While $z > 0$:
  - increment $n_{\text{iter}} = n_{\text{iter}} + 1$
  - draw at random a realization of $\boldsymbol{x}$, denoted by $\boldsymbol{x}^\star$
  - draw at random a realization of $Y_j(\boldsymbol{x}^\star)$, denoted by $y^\star$
  - If $y^\star < z$, actualize : $z = y^\star$, $M^j = M^j + 1$ if $y^\star > 0$, $Y_j(\omega) = Y_j(\omega) \mid Y_j(\boldsymbol{x}^\star) = y^\star$, $\widehat{\mathcal{X}}^j = \widehat{\mathcal{X}}^j \cup \{\boldsymbol{x}^\star\}$, $\widehat{\mathcal{Y}}^j = \widehat{\mathcal{Y}}^j \cup \{y^\star\}$.
- Compute $\widetilde{p}_j := \left(1 - \frac{1}{Q}\right)^{M^j}$.

$\Rightarrow$ By taking the $j^\star(\alpha,\beta)^{\text{th}}$ biggest value among $\widetilde{p}_1, \ldots, \widetilde{p}_m$, we obtain a value with more than $1-\beta$ chance of being larger than the $1-\alpha$ quantile of P_f^Y.

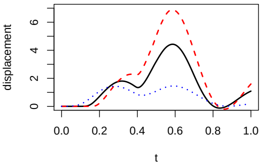
Coupling GPR and point process-based approaches
Back to the introduction example



(a) Real tank



(b) Time evolution of the pressure



(c) Time evolution of the displacement u

FIGURE: Pressure tank under dynamic pressure

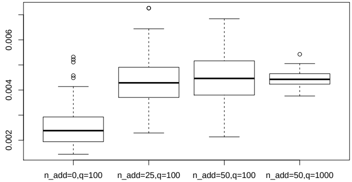
$$p_f := \mathbb{P}_{\boldsymbol{X}}\left(\max_{\text{time,space}} u(\boldsymbol{X}) > s\right).$$

$$\boldsymbol{X} = \{\text{geometry and material uncertainties}\}.$$

Coupling GPR and point process-based approaches

Results

- We first compute the value of y in $\ell = 50$ points uniformly chosen in the input space, and construct the GPR Y of y. **None** of these values of y was over s.
- $m = 100$ estimators of P_f^Y were computed using Y.
- There are **two sources** for the dispersion : the variability related to Y (which can be reduced by adding n_{add} new code evaluations) and the variability related to the estimator (which can be reduced by increasing Q).



Comparison of the dispersions obtained on the estimates of p_f as a function of the number of points added n_{add} and the number of Poisson processes Q.

Coupling GPR and point process-based approaches

Application à la cuve EPURE

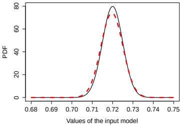
Applied on the mean function of Y, sensitivity and robustness analyses can also be carried out.

	$\widehat{s}_i(\%)$	$\widehat{t}_i(\%)$	$\widehat{\varsigma}_i^{\delta}(\%)$	$\widehat{\tau}_i^{\delta}(\%)$
i=1	[0 ;2]	[44 ;89]	[0 ;4]	[1 ;7]
i=2	[0 ;2]	[50 ;91]	[2 ;9]	[3 ;14]
i=3	[0 ;0]	[47 ;89]	[3 ;9]	[5 ;15]
i=4	[0 ;1]	[64 ;93]	[5 ;8]	[12 ;16]
i=5	[4 ;13]	[88 ;99]	[45 ;64]	[57 ;76]
i=6	[0 ;0]	[45 ;89]	[2 ;10]	[3 ;17]
i=7	[0 ;0]	[43 ;89]	[1 ;5]	[1 ;9]
i=8	[0 ;0]	[43 ;90]	[0 ;2]	[1 ;8]

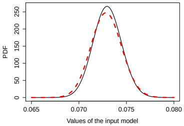
TABLE: The values in square brackets are 95% intervals, incorporating the uncertainty due to the metamodel and the approximate nature of the probability, but based on the same metamodel. $p^{\delta} = 0.077$ so that $p^{\delta} \approx 2p_f$.

$\Rightarrow$ the threshold is exceeded for configurations requiring very specific combinations of **all** parameters.

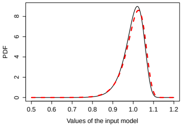
Coupling GPR and point process-based approaches
Application à la cuve EPURE



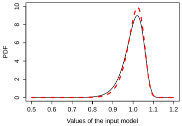
(a) x_1



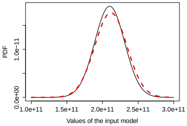
(b) x_2



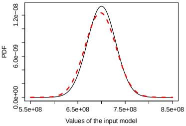
(c) x_3



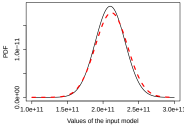
(d) x_4



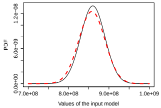
(e) x_5



(f) x_6



(g) x_7



(h) x_8

FIGURE: Comparison between the initial PDFs (in black) and the perturbed PDFs (in dotted red) leading to the multiplication of p_f by a factor 2.

Outline of the presentation

Conclusions

- To guarantee by simulation, it is necessary to evaluate a **double probability** : the first associated with a risk, the second to attribute a confidence to the estimator.
- The guarantee problem can be expressed on the quantile as well as on the probability of exceeding a threshold.
- The lower the risk, and the more the guarantee requires a large number of calls to the code $\Rightarrow$ one is often obliged to use **metamodels** for low probabilities, **whose uncertainty must also be propagated in the model**.

Conclusions

- The guarantee is **conditioned** by the fact :
  - that one has confidence in the theory of probabilities to model the uncertain,
  - that one knows how to generate random numbers,
  - that the sources of uncertainty have been well identified and modelled.
- In order to be more conservative with regard to uncertainties on the laws of the inputs, we note several recent developments (notably at EDF) of so-called "robust Bayes" approaches, which seek to solve problems of the type :

$$P_f^* = \arg \max_{f_y, f_{\boldsymbol{x}} \in \mathbb{F}_y \times \mathbb{F}_x} \mathbb{P}(y(\boldsymbol{x} > S) \mid \boldsymbol{x} \sim f_{\boldsymbol{x}},\ y(\boldsymbol{x})|\boldsymbol{x} \sim f_y),$$

  where $\mathbb{F}_y$ and $\mathbb{F}_x$ are sets of laws constrained by the "true" knowledge about $\boldsymbol{x}$...

Conclusion

More details on the **moving particle algorithm** can be found here :

- C. Walter, Using Poisson processes for rare event simulation.
- G. Defaux, G. Perrin, C. Walter, Point process-based approaches for the reliability analysis of systems with multiple failure modes, ICOSSAR 2017.

More details on the **robust and sensitivity analyses** can be found here :

- G. Perrin, C. Soize, N. Ouhbi, Data-driven kernel representations for sampling with an unknown block dependence structure under correlation constraints, CSDA, 2018.
- G. Perrin, G. Defaux, Efficient Evaluation of Reliability-Oriented Sensitivity Indices, Journal of Scientific Computing, 2019.

More details on the **coupling between GPR and MV** can be found here :

- G. Perrin, Point process-based approaches for the reliability analysis of systems modeled by costly simulators, RESS, 2021.

Thank you for your attention.