Point process-based approaches for the reliability analysis of systems modeled by costly simulators

G. Perrin[1]

RESIM 2021 | May 19^{th} 2021

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

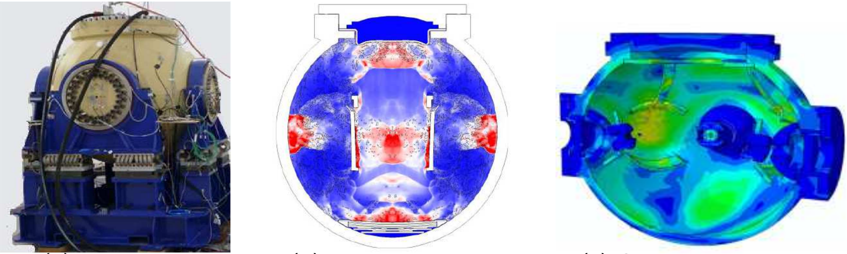
Introduction

- 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, how to bound this failure probability with a prescribed confidence ?

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

General framework

Notations

- $\mathcal{S} \leftrightarrow$ system of interest,
- $\boldsymbol{x} \in \mathbb{X} \subset \mathbb{R}^D \leftrightarrow$ system characteristics (dimensions, boundary conditions, material properties...),
- $\boldsymbol{x} \mapsto y(\boldsymbol{x}) \in \mathbb{R} \leftrightarrow$ quantity of interest for the monitoring of $\mathcal{S}$,
- $\mathcal{F} = \{\boldsymbol{x} \in \mathbb{X} \mid y(\boldsymbol{x}) < 0\} \leftrightarrow$ system's failure domain.

Assumption

$\boldsymbol{x}$ is not perfectly known $\Rightarrow$ it is modeled by a r.v. $\boldsymbol{X}$ with known PDF $f_{\boldsymbol{X}}$.

$\Rightarrow$ $p_f := \mathbb{P}_{\boldsymbol{X}}(y(\boldsymbol{X}) < 0) = \int_{\mathcal{F}} f_{\boldsymbol{X}}(\boldsymbol{x}) d\boldsymbol{x}$ $\leftrightarrow$ system failure probability of interest.

Need for surrogate models

$$p_f := \mathbb{P}_{\boldsymbol{X}}(y(\boldsymbol{X}) < 0) = \int_{\mathcal{F}} f_{\boldsymbol{X}}(\boldsymbol{x}) d\boldsymbol{x}.$$

- $y \leftrightarrow$ output of a numerically **expensive deterministic "black box"** : for each $\boldsymbol{x}$, $y(\boldsymbol{x})$ is unique, and can be calculated by using a simulator that can take a long time to evaluate.

⇒ In this type of configuration, 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 $Y \sim \mathrm{GP}(\mu, \Sigma)$.
- Under that formalism, $p_f = \mathbb{P}_{\boldsymbol{X}}\left(Y(\boldsymbol{X}) < 0 \mid Y = y\right)$.

⇒ p_f is a particular realization of the random variable :

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

Surrogate modeling and reliability analysis

Assuming that Y is a good approximation of y, p_f can then be approximated by the mean value $\widehat{p}_f$ of P_f^Y (or possibly by $\widetilde{p}_f := \mathbb{P}_{\boldsymbol{X}}\left(\mu(\boldsymbol{X}) < 0\right)$) :

$$\widehat{p}_f = \mathbb{E}_Y\left[P_f^Y\right] = \mathbb{E}_{\boldsymbol{X}}\left[\Phi\left(-\frac{\mu(\boldsymbol{X})}{\sqrt{\Sigma(\boldsymbol{X},\boldsymbol{X})}}\right)\right], \quad \Phi(u) = \int_{-\infty}^{u} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{v^2}{2}\right) dv.$$

Sampling techniques can finally be used to estimate $\widehat{p}_f$ (or $\widetilde{p}_f$) without requiring any additional evaluation of expensive function y.

Surrogate modeling and reliability analysis

Assuming that Y is a good approximation of y, p_f can then be approximated by the mean value $\widehat{p}_f$ of P_f^Y (or possibly by $\widetilde{p}_f := \mathbb{P}_{\boldsymbol{X}}(\mu(\boldsymbol{X}) < 0))$:

$$\widehat{p}_f = \mathbb{E}_Y\left[P_f^Y\right] = \mathbb{E}_{\boldsymbol{X}}\left[\Phi\left(-\frac{\mu(\boldsymbol{X})}{\sqrt{\Sigma(\boldsymbol{X},\boldsymbol{X})}}\right)\right], \quad \Phi(u) = \int_{-\infty}^{u} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{v^2}{2}\right) dv.$$

Sampling techniques can finally be used to estimate $\widehat{p}_f$ (or $\widetilde{p}_f$) without requiring any additional evaluation of expensive function y.

However, replacing true function y by an accurate surrogate can still lead to an inaccurate estimation of p_f...

Surrogate modeling and reliability analysis

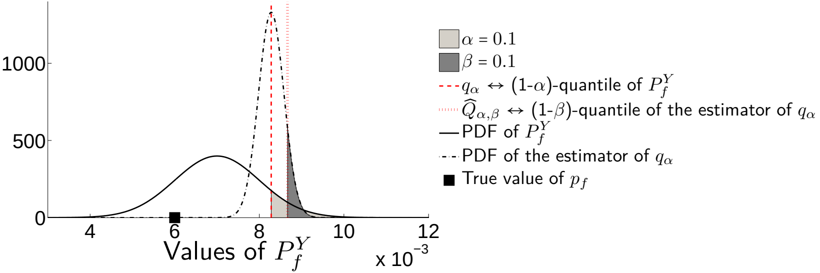
- To correctly anticipate the risks of deterioration of the system, we propose to work on the construction of **confidence bounds** to failure probability estimates.
- Instead of working on the estimation of the mean value of P_f^Y, we would like to construct a robust estimator $\widehat{Q}_{\alpha,\beta}$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

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

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

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.

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,
  2. the Gaussian process regression formalism,
  3. a particular Marked Poisson Process.

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,
  2. the Gaussian process regression formalism,
  3. a particular Marked Poisson Process.
- **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.

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,
  2. the Gaussian process regression formalism,
  3. a particular Marked Poisson Process.
- **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.

Outline

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

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.

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 :

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

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 $\widehat{Q}_{\alpha,\beta}$.

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).
$\Rightarrow$ **another estimator is needed !**

Choice of the estimator (2/2)

- If $P_1, \dots, P_q$ are q independent copies of a Poisson process $P(-\log(\mathbb{P}_{\boldsymbol{X}}(Y(\boldsymbol{X};\omega) < 0)))$,

$$\widehat{P}_f^{Y,\mathcal{X}_n}(\omega) := \left(1 - \frac{1}{q}\right)^{\sum_{k=1}^q P_k}$$

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,
  - $Y(\omega)$ only needs to be projected in $\mathbb{E}\left[\sum_{k=1}^q P_k\right] = -q\log(P_f^Y(\omega))$ points in average ($\ll 100/P_f^Y(\omega)$ for the former MC approach).

- For $Z := Y(\boldsymbol{X};\omega)$, we can then notice that

$$P(-\log(\mathbb{P}_{\boldsymbol{X}}(Y(\boldsymbol{X};\omega) < 0))) = \sup\{i;\ Z_i \geq 0\},$$

$$Z_0 = +\infty,\ \mathbb{P}(Z_{i+1} \leq z \mid Z_i) = \mathbb{P}(Z \leq z \mid Z \leq Z_i),$$

such that realizations of $\widehat{P}_f^{Y,\mathcal{X}_n}$ can be obtained by launching in parallel on the same instance of the random process Y several draws of $\{Z_i, i \geq 0\}$.

Outline

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 a number of decreasing walks 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 $\boldsymbol{X}(\omega_1), \dots, \boldsymbol{X}(\omega_q)$
  - Sample one realization of the Gaussian vector $(Y(\boldsymbol{X}(\omega_1)), \dots, Y(\boldsymbol{X}(\omega_q)))$, noted $(y_1, \dots, y_q)$
  - Define $Y_j(\omega) := Y \mid Y(\boldsymbol{X}(\omega_k)) = y_k, \ 1 \leq k \leq q$
  - Set $n_{\text{iter}} = 0$, $\widehat{\mathcal{X}}^j = \{\boldsymbol{X}(\omega_1), \dots, \boldsymbol{X}(\omega_q)\}$, $\widehat{\mathcal{Y}}^j = \{y_1, \dots, y_q\}$.

Practical implementation

Iteration

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

- For $1 \leq k \leq q$:
  - Set $z = y_k$, $P_k^j = 0$
  - 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 : $z = y^\star$, $P_k^j = P_k^j + 1$ $Y_j(\omega) = Y_j(\omega) \mid Y_j(\boldsymbol{x}^\star) = y^\star$, $\widehat{\mathcal{X}}^j = \widehat{\mathcal{X}}^j \cup \{\boldsymbol{x}^\star\}$, $\widehat{\mathcal{Y}}^j = \widehat{\mathcal{Y}}^j \cup \{y^\star\}$.
- Compute $\widetilde{p}_j := \left(1 - \frac{1}{q}\right)^{\sum_{k=1}^q P_k^j}$.

Practical implementation

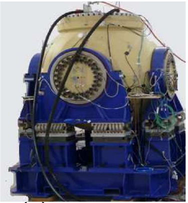
Iteration

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

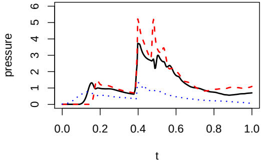
- For $1 \leq k \leq q$:
  - Set $z = y_k$, $P_k^j = 0$
  - While $z > 0$:
    - increment $n_{\text{iter}} = n_{\text{iter}} + 1$
    - draw at random a realization of $\boldsymbol{X}$, denoted by $\boldsymbol{x}^\star$
    - draw at random a realization of $Y_j(\boldsymbol{x}^\star)$, denoted by $y^\star$
    - If $y^\star < z$, actualize : $z = y^\star$, $P_k^j = P_k^j + 1$ $Y_j(\omega) = Y_j(\omega) \mid Y_j(\boldsymbol{x}^\star) = y^\star$, $\widehat{\mathcal{X}}^j = \widehat{\mathcal{X}}^j \cup \{\boldsymbol{x}^\star\}$, $\widehat{\mathcal{Y}}^j = \widehat{\mathcal{Y}}^j \cup \{y^\star\}$.
- Compute $\widetilde{p}_j := \left(1 - \frac{1}{q}\right)^{\sum_{k=1}^q P_k^j}$.

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

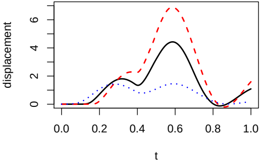
Back to the introduction example



(a) Real tank



(b) Time evolution of the pressure



(c) Time evolution of the displacement

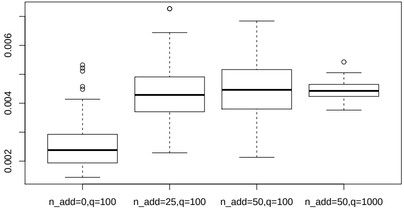
FIGURE: Pressure tank under dynamic pressure

$$p_f := \mathbb{P}_{\boldsymbol{X}}(\max_{t,\boldsymbol{z}} u(t,\boldsymbol{z}) > s).$$

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

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.

Outline

Conclusions and prospects

- This presentation introduces a formalism for estimating probabilities of failure.
- This approach is based on : GPR, order statistics, a marked Poisson process.
- In order to ensure the security of systems of interest, it is proposed to focus on the estimation of quantiles rather than the mean.
- One of the objectives of the method is to avoid forgetting pathological configurations in the risk analysis.
- A sequential enrichment criterion particularly dedicated to the estimation method can be found in :

G. Perrin. Point process-based approaches for the reliability analysis of systems modeled by costly simulators. Reliability Engineering and System Safety, Elsevier, In press. <hal-03228196>.

Thank you for your attention.