

DE LA RECHERCHE À L'INDUSTRIE



Garantir par la simulation : comment les incertitudes créent la confiance

G. Perrin

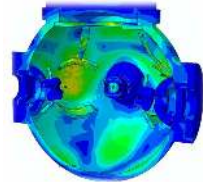
CEA/DAM/DIF, Arpajon, France

ETICS 2018 - Roscoff |

4 Juin 2018

- 1 Introduction
- 2 Garantir une probabilité de dépassement de seuil
- 3 Garantir un quantile
- 4 Sensibilité aux choix de modélisation
- 5 Conclusion

La bonne tenue d'une cuve de confinement pour des études en dynamique rapide



A partir de mesures non-destructives, une masse de radionucléides dans des fûts de déchets mal-connus à entreposer



Quelques définitions de "garantir":

- 1 Assurer, sous sa **responsabilité**, l'exécution de quelque chose dans des conditions préalablement définies.
- 2 **Se porter officiellement garant** des qualités et du bon fonctionnement de quelque chose, le placer sous garantie et **s'engager** à remédier à tout défaut, toute panne etc. pouvant se présenter pendant un certain temps.
- 3 Donner quelque chose comme sûr, vrai, certain, sous sa **responsabilité**; assurer, répondre de, certifier.
- 4 Déclarer quelque chose conforme à certaines normes de qualité ou **assurer** son authenticité.

Qui serait assez fou pour s'engager autant sur les seules bases d'un résultat de simulation?

... et encore, on ne vit pas au temps d'Hammurabi...

Quelques définitions de "garantir":

- 1 Assurer, sous sa responsabilité, l'exécution de quelque chose **dans des conditions préalablement définies.**
 - 2 Se porter officiellement garant **des qualités et du bon fonctionnement de quelque chose**, le placer sous garantie et s'engager à remédier à tout défaut, toute panne etc. pouvant se présenter **pendant un certain temps.**
 - 3 Donner quelque chose comme sûr, vrai, certain, sous sa responsabilité; assurer, répondre de, certifier.
 - 4 Déclarer quelque chose conforme **à certaines normes de qualité** ou assurer son authenticité.
- La simulation n'est qu'un **outil d'aide** à la garantie...
 - Par la simulation, on ne peut s'engager que **sous conditions...**

Dans cette présentation, on note:

- \mathcal{S} le système d'intérêt à garantir, que l'on suppose pouvoir être décrit par un vecteur $\boldsymbol{x} \in \mathbb{X}$.
- $y(\boldsymbol{x}) \in \mathbb{Y}$ la quantité d'intérêt (Qol) permettant d'analyser le bon fonctionnement de \mathcal{S} .

Le vecteur \boldsymbol{x} est modélisé par un vecteur aléatoire de loi (connue) $f_{\boldsymbol{x}}$ pour traduire les incertitudes sur les entrées (tolérance de fabrication, méconnaissance).

De même, à valeur de \boldsymbol{x} donnée, $y(\boldsymbol{x})|\boldsymbol{x}$ est modélisé par une variable aléatoire de loi (connue) $f_{y(\boldsymbol{x})|\boldsymbol{x}}$ pour traduire les incertitudes sur la sortie (erreur de modèle, erreur de précision, erreur sur les paramètres...).

La loi de $y(\boldsymbol{x})$, $f_{y(\boldsymbol{x})}(y) = \int_{\mathbb{X}} f_{y(\boldsymbol{x})|\boldsymbol{x}=\boldsymbol{z}}(y) f_{\boldsymbol{x}}(\boldsymbol{z}) d\boldsymbol{z}$, est bien définie en théorie mais généralement inaccessible en pratique.

- D'un point de vue **déterministe**, garantir le bon fonctionnement du système peut se ramener à garantir que $y(\mathbf{x}) \leq S$, où la valeur de S est une donnée du problème (fournie par un client, une norme...).
- Une fois les incertitudes considérées, cela revient à garantir que $\mathbb{P}(y(\mathbf{x}) > S) = 0$, ce qu'il n'est quasiment jamais possible de faire en pratique.
- Pour un risque $0 \leq \alpha \leq 1$, une version **probabiliste** plus *réaliste* de cette garantie consiste à démontrer que :

$$P_f := \mathbb{P}(y(\mathbf{x}) > S) \leq \alpha.$$

- Si \hat{P}_f définit un estimateur statistique de P_f , la garantie de bon fonctionnement du système demande la définition d'une transformation \mathcal{T} telle que:

$$\mathbb{P}(P_f \leq \mathcal{T}(\hat{P}_f)) \geq \beta,$$

où β est un niveau de confiance, avec $\mathcal{T}(\hat{P}_f)$ à comparer à α .

La garantie d'un système par la simulation (physique, chimique, mécanique, financier...) nécessite ainsi l'introduction de plusieurs grandeurs:

- $\mathbf{x} = (x_1, \dots, x_d) \leftrightarrow$ paramétrage du système, que l'on modélise comme un **vecteur aléatoire**,
- $y \leftrightarrow$ quantité d'intérêt pour la surveillance du bon fonctionnement du système considéré, qui peut elle-même être aléatoire pour une valeur de \mathbf{x} donnée,
- $S \leftrightarrow$ seuil à ne pas dépasser pour y ,
- $\alpha \leftrightarrow$ risque "acceptable" (de dépassement de seuil),
- $\beta \leftrightarrow$ niveau de confiance (sur l'estimateur statistique).

$$\mathbb{P}(P_f \leq \mathcal{T}(\widehat{P}_f)) \geq \beta, \quad \mathcal{T}(\widehat{P}_f) \sim \alpha.$$

- Pour "ne pas prendre de risque", un certain nombre de systèmes sont *a priori* conçus "loin" du seuil S . Cela suppose que la valeur réelle (bien qu'inconnue) de $P_f = \mathbb{P}(y(\mathbf{x}) > S)$ est très faible (entre 10^{-3} et 10^{-6} par exemple). Des méthodes spécifiques sont ainsi requises pour proposer des estimateurs efficaces en temps de calcul de telles probabilités (parties 2 et 3).
- La valeur de P_f est complètement contrôlée par le choix/l'identification des distributions de \mathbf{x} et de $y(\mathbf{x})|\mathbf{x}$. Une fois la valeur de P_f estimée, il est souvent intéressant de chercher à identifier les sources d'incertitudes les plus influentes sur cette probabilité, et de contrôler la robustesse du résultat par rapport à de légères fluctuations de ces distributions (partie 4).

La **théorie des valeurs extrêmes** peut être perçue comme un post-traitement "paramétrique" d'un très grand nombre d'évaluations de y en des points x tirés aléatoirement et indépendamment, pour l'évaluation de P_f . Les méthodes qui vont être présentées dans la suite s'en distinguent par le fait que l'on suppose ici:

- qu'une évaluation de $y(x)$ est relativement "**coûteuse**" (financièrement pour une expérience ou numériquement pour une simulation),
- qu'à l'état initial, **aucune** évaluation de y n'a été effectuée.

⇒ on cherche plutôt à mettre au point une stratégie itérative pour l'évaluation du code en des points bien choisis pour maximiser la précision d'estimation de P_f à nombre d'évaluation minimal.

- 1 Introduction
- 2 Garantir une probabilité de dépassement de seuil
- 3 Garantir un quantile
- 4 Sensibilité aux choix de modélisation
- 5 Conclusion

$$P_f := \mathbb{P}(y(\mathbf{x}) > S) = \int_S^{+\infty} f_{y(\mathbf{x})}(y) dy = \int_{\mathbb{R} \times \mathbb{X}} \mathbf{1}_{y > S} f_{y(\mathbf{x})|\mathbf{x}}(y) f_{\mathbf{x}}(\mathbf{x}) d\mathbf{x} dy$$

- L'évaluation de P_f par une intégration multidimensionnelle par quadrature est très rapidement non envisageable en fonction de la dimension de \mathbf{x} .
- Les méthodes d'intégration par échantillonnage sont alors souvent préférées.
- L'idée est alors de définir un enchainement **dépendant** (au sens où les résultats passés sont utilisés pour définir les nouveaux points d'évaluation) de points \mathbf{x} , et donc d'évaluations de y , de taille **minimum**, permettant la **meilleure** estimation de P_f .

L'approche Monte Carlo est basée sur l'écriture suivante de P_f :

$$P_f := \mathbb{P}(y(\mathbf{x}) > S) = \mathbb{E} [1_{y(\mathbf{x}) > S}].$$

Ainsi, pour évaluer P_f , on peut tirer N réalisations iid de \mathbf{x} , que l'on note $\{\mathbf{x}_1, \dots, \mathbf{x}_N\}$, et évaluer y en chacun de ces points: $y_n := y(\mathbf{x}_n)$. A partir de ces tirages, on peut écrire:

$$P_f \approx \hat{P}_f := \frac{1}{N} \sum_{n=1}^N 1_{y_n > S}.$$

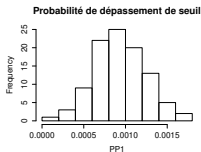
Propriétés statistiques de l'estimateur

- $\mathbb{E} [\hat{P}_f] = P_f.$
- $\text{Var}(\hat{P}_f) = \frac{P_f(1-P_f)}{N}.$
- $\hat{P}_f \xrightarrow{N \rightarrow +\infty} \mathcal{N}(P_f, \frac{P_f(1-P_f)}{N}).$

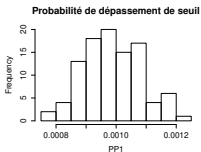
Le lien entre N et le carré du coefficient de variation est ainsi donné par:

$$\delta_{MC}^2 := \frac{\text{Var}(\hat{P}_f)}{\mathbb{E}[\hat{P}_f]^2} = \frac{1 - P_f}{P_f N} \Leftrightarrow N = \frac{1 - P_f}{P_f \delta_{MC}^2}.$$

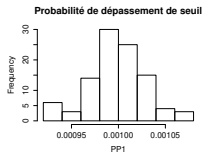
Illustration pour l'estimation de $\mathbb{P}(X > 3.09) = 10^{-3}$ pour $X \sim \mathcal{N}(0, 1)$.



(a) $N = 10^4$



(b) $N = 10^5$



(c) $N = 10^6$

- (+) Très facile à implémenter.
- (+) Gère sans problème le caractère hiérarchique de l'incertitude.
- (-) Demande potentiellement un très grand nombre d'appels au code pour assurer une bonne précision.

$$\widehat{P}_f \xrightarrow{N \rightarrow +\infty} \mathcal{N} \left(P_f, \frac{P_f(1 - P_f)}{N} \right).$$

En notant c_β le quantile d'ordre β de la loi gaussienne centrée réduite, et

$$\mathcal{T}(\widehat{P}_f) = \widehat{P}_f + c_\beta \sqrt{\frac{\widehat{P}_f(1 - \widehat{P}_f)}{N}}:$$

$$\mathbb{P} \left(P_f \leq \mathcal{T}(\widehat{P}_f) \right) \approx \beta = \int_{-\infty}^{c_\beta} \frac{1}{\sqrt{2\pi}} \exp \left(-\frac{x^2}{2} \right) dx.$$

Par exemple, $c_\beta = 0$ pour $\beta = 50\%$, $c_\beta \approx 1.28$ pour $\beta = 90\%$ et $c_\beta \approx 2.33$ pour $\beta = 99\%$.

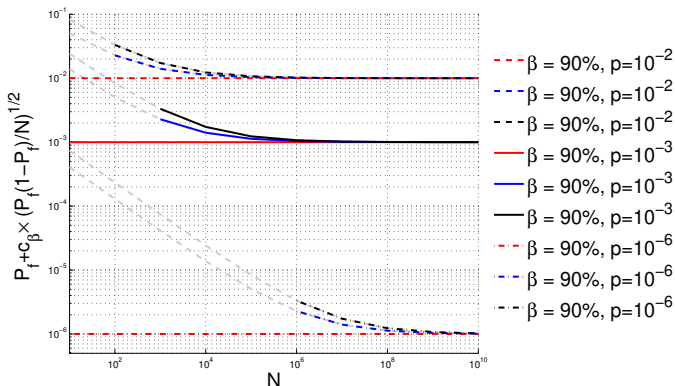


Figure: Valeurs de α qu'il serait possible de garantir pour trois valeurs de P_f ($p = 10^{-2}, 10^{-3}, 10^{-6}$), deux valeurs de β (90% et 99% en fonction du nombre d'appels au code (en supposant que l'on obtient au moins un calcul défaillant...)).

Afin de réduire ce nombre important d'évaluations, il est intéressant de noter que si $S_0 = -\infty$ et $S_1 < S_2 < \dots < S_M = S$:

$$\mathbb{P}(y(\mathbf{x}) \geq S) = \prod_{m=1}^M P_m, \quad P_m := \mathbb{P}(y(\mathbf{x}) \geq S_m | y(\mathbf{x}) \geq S_{m-1}).$$

- On peut ainsi définir M estimateur MC indépendants pour chaque probabilité P_m , que l'on nomme \hat{P}_m .
- $\hat{P}_f^C := \prod_{m=1}^M \hat{P}_m$ définit un nouvel estimateur de P_f .

- Se pose alors le problème du choix de M , des seuils S_1, \dots, S_M , et du nombre de tirages à attribuer à l'estimation de chaque probabilité P_m .
- On peut montrer que pour une valeur de M fixée, le coefficient de variation de l'estimateur est minimal pour $P_1 = \dots = P_M = \mathbb{P}(y(\mathbf{x}) \geq S)^{1/M}$, $N_1 = \dots = N_M = N/M$.
- Dans ce cas, on déduit la valeur du coefficient de variation:

$$\delta^2 = (\delta_0^2 + 1)^M - 1, \quad \delta_0^2 = (P_f^{-1/M} - 1)M/N.$$

- On peut minimiser ce coefficient de variation (en prenant M et N continus), et on obtient $M \approx -0.63 \log(P_f) \leftrightarrow P_m \approx 0.2032$.

Commenter les limites pratiques d'un tel tirage conditionné "optimal".

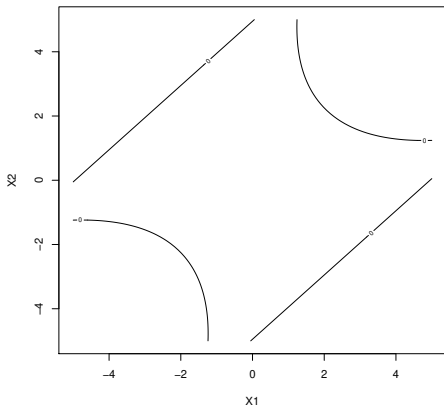
Face à un problème réel, dont on ne connaît pas la probabilité de défaillance, deux méthodes inspirées du tirage conditionné "optimal" sont proposées:

- les approches dites "**subset simulation**" : on choisit une valeur de p_0 (souvent 0.1) et de N_0 (entre 100 et 1000), puis on rééchantillonne itérativement les $(1 - p_0)N_0$ plus faibles valeurs pour être plus grand que la p_0N_0 plus haute valeur. On s'arrête quand cette p_0N_0 plus haute valeur dépasse S .
- les approches dites "**moving particle**" : on choisit une valeur N_0 de taille d'échantillon initiale. On "déplace" itérativement la "particle" associée à la valeur minimale de la quantité d'intérêt conditionnement à sa propre valeur jusqu'à ce que toutes les particules dépassent S .

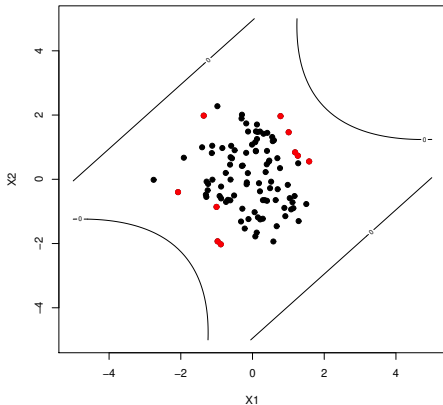
Illustration sur la fonction de Waarts :

$$b_1 = 3 + (u_1 - u_2)^2/10 - \text{sign}(u_1 + u_2)(u_1 + u_2)/\sqrt{2}, b_2 = \text{sign}(u_2 - u_1)(u_1 - u_2) + 7/\sqrt{2}, f(u) = \min(b_1, b_2).$$

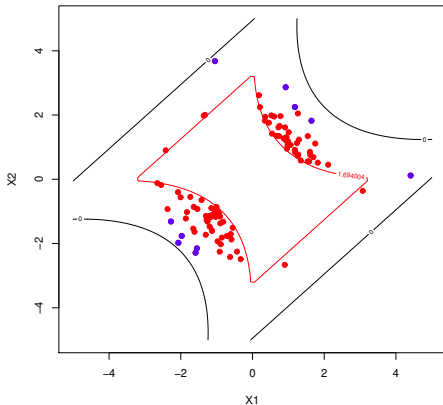
$p_0 = 0.1$, $N_0 = 100$. Ligne noire : $y(\mathbf{x}) = 0$.



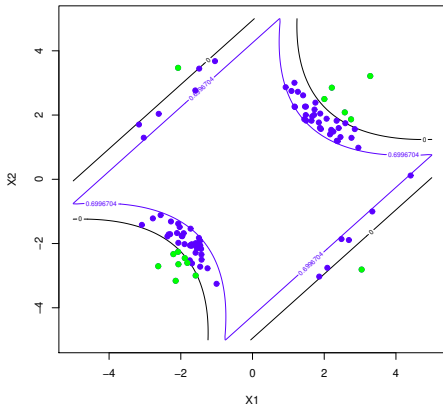
$p_0 = 0.1$, $N_0 = 100$. Rouge : $p_0 N_0 = 10$ points $> S_1$.



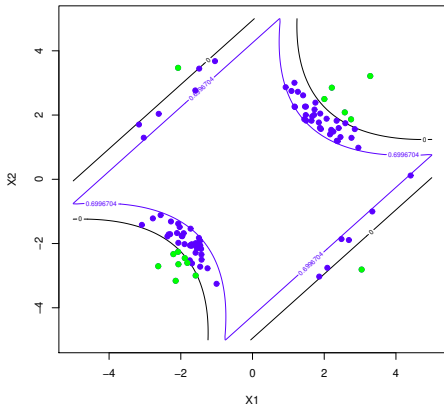
$p_0 = 0.1$, $N_0 = 100$. Bleu : $p_0 N_0 = 10$ points $> S_2$.



$p_0 = 0.1$, $N_0 = 100$. Vert : 15 points au dessus de S .

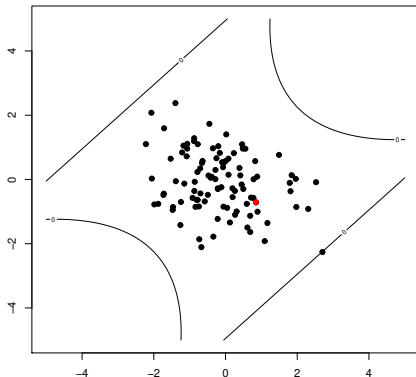


$p_0 = 0.1$, $N_0 = 100$. Vert : 15 points au dessus de S .

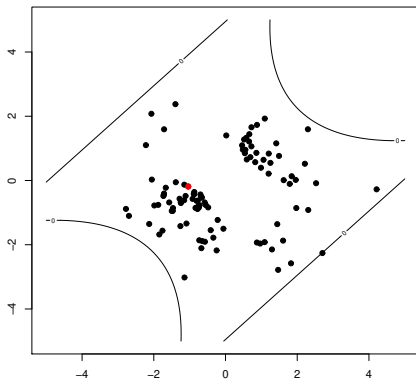


$$P_f \approx p_0^2 \times \frac{\#\text{points} > S}{N_0} \approx 0.0015.$$

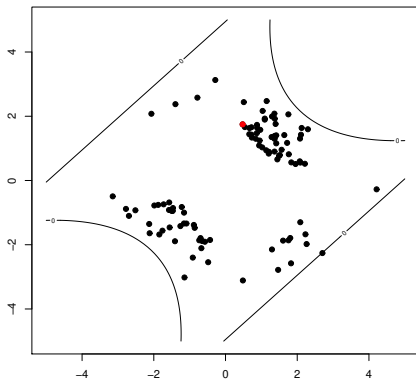
$N_0 = 100$. Rouge: point minimum lors du tirage initial (0 déplacements).



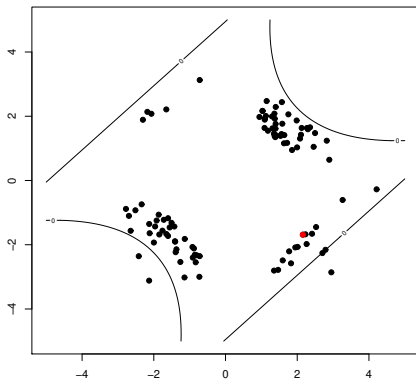
$N_0 = 100$. Rouge: point minimum après 100 déplacements.



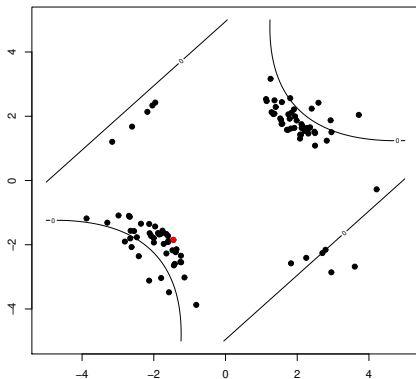
$N_0 = 100$. Rouge: point minimum après 200 déplacements.



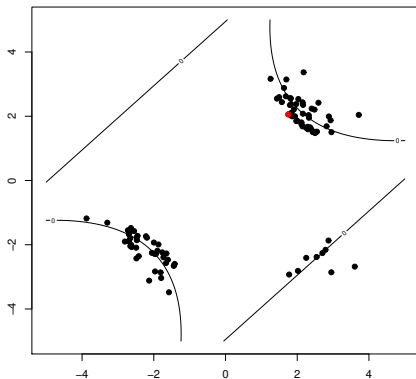
$N_0 = 100$. Rouge: point minimum après 300 déplacements.



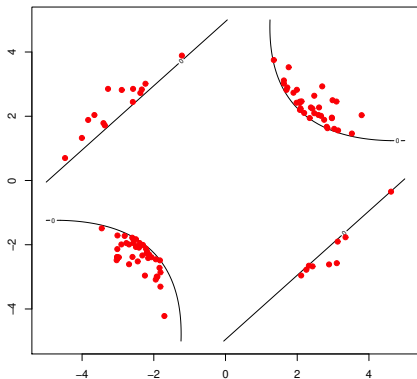
$N_0 = 100$. Rouge: point minimum après 400 déplacements.



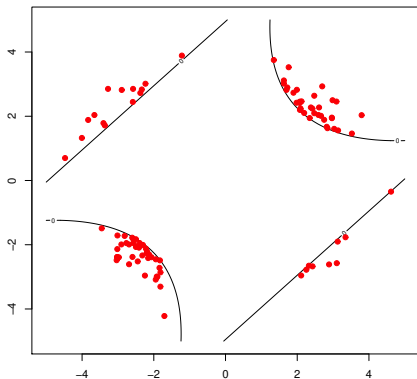
$N_0 = 100$. Rouge: point minimum après 500 déplacements.



$N_0 = 100$. Après $M=601$ déplacements, toutes les particules sont $> S$.



$N_0 = 100$. Après $M=601$ déplacements, toutes les particules sont $> S$.



$$P_f \approx \left(1 - \frac{1}{N_0}\right)^{M-1} \approx 0.0024.$$

- (+) demande un nombre a priori réduit d'évaluations de y .
- (-) plus difficile à mettre en oeuvre (nécessité de pouvoir tirer des échantillons conditionnellement à un seuil donné sur $y(\mathbf{x})$).
- ⇒ passer par des approches de génération itérative de type Markov Chain Monte Carlo (MCMC).
- ⇒ cela se traduit en pratique par la nécessité d'effectuer T évaluations du code pour chaque déplacement de particules.
- ⇒ En notant M le nombre d'évaluations du code inférieures à S , le nombre total d'évaluations est alors égal à

$$N = N_0 + MT.$$

On peut noter que $M \sim \text{Poisson}(-N_0 \log(P_f))$. Ainsi, $\hat{P}_f = (1 - 1/N_0)^M$ est un nouvel estimateur non biaisé de P_f . En utilisant l'approximation gaussienne de la loi de Poisson pour des valeurs de $-N_0 \log(P_f)$ pas trop petites (≥ 5), on déduit $N \approx N_0(1 - T \log(p))$ et

$$\begin{aligned}\log(\hat{P}_f) &\approx \mathcal{N}(-N_0 \log(P_f) \log(1 - 1/N_0), -N_0 \log(P_f) \log(1 - 1/N_0)^2) \\ &\approx \mathcal{N}(\log(P_f), -\log(P_f)/N_0).\end{aligned}$$

En notant c_β le quantile d'ordre β de la loi gaussienne centrée réduite, et

$$\mathcal{T}(\hat{P}_f) = \hat{P}_f \exp\left(c_\beta \sqrt{\frac{-\log(\hat{P}_f)}{N_0}}\right), \text{ on déduit:}$$

$$\mathbb{P}\left(P_f \leq \mathcal{T}(\hat{P}_f)\right) \approx \beta.$$

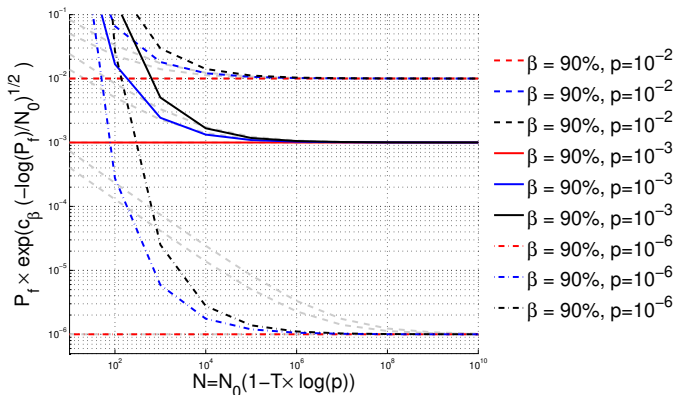


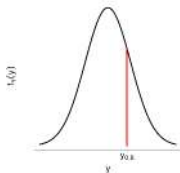
Figure: Valeurs de α qu'il serait possible de garantir pour trois valeurs de P_f ($p = 10^{-2}$, 10^{-3} , 10^{-6}), deux valeurs de β (90% et 99% en fonction du nombre d'appels au code $N \approx N_0(1 - T \log(p))$, avec $T = 10$ et $N_0 = 100$ (les pointillés grisés correspondent à l'approche MCS).

- 1 Introduction
- 2 Garantir une probabilité de dépassement de seuil
- 3 Garantir un quantile
- 4 Sensibilité aux choix de modélisation
- 5 Conclusion

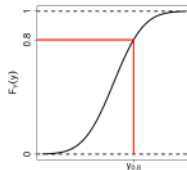
- Plutôt que de parler de probabilité d'"accident" ou de "défaillance", un certain nombre d'industriels préfèrent reformuler le problème sous la forme de la garantie d'un quantile.
- Dans ce cas, on spécifie non plus un seuil pour $y(\mathbf{x})$ mais un risque de dépassement, et on se sert du quantile déduit pour dimensionner l'objet considéré.
- Pour rappel, le quantile q_α de niveau $0 \leq \alpha \leq 1$ associé à $y(\mathbf{x})$ est défini par:

$$\mathbb{P}(y(\mathbf{x}) \leq q_\alpha) = \alpha \leftrightarrow q_\alpha = \arg \min_q |F_{y(\mathbf{x})}(q) - \alpha|.$$

Densité de probabilité de Y (pdf)



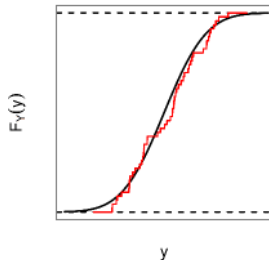
Fonction de répartition de Y (cdf)



- Soient $\{\mathbf{x}^{(n)}\}_{n=1, \dots, N}$ un N -échantillon de réalisations indépendantes de $\mathbf{x} \in \mathbb{R}^d$, et $\{Y_n = y(\mathbf{x}^{(n)})\}_{n=1, \dots, N}$ les simulations correspondantes.
- On note $\{Y_n^*\}_{n=1, \dots, N}$ l'échantillon ordonné : $Y_1^* \leq \dots \leq Y_N^*$.
- L'estimateur Monte Carlo de la CDF de $y(\mathbf{x})$, notée $F_{y(\mathbf{x})}$, s'écrit :

$$\hat{F}_N(y) = \frac{1}{n} \sum_{n=1}^N 1_{Y_n \leq y}.$$

$\hat{F}_N(y) \leftrightarrow$ pourcentage d'observations de l'échantillon inférieures ou égales à y .



- Estimateur sans biais : $\mathbb{E}[\hat{F}_N(y)] = F_{y(\mathbf{x})}(y)$
- $\hat{F}_N(y) \xrightarrow{N \rightarrow \infty} F_{y(\mathbf{x})}(y)$ avec probabilité 1 (Th. de Glivenko-Cantelli)

- Estimateur Monte Carlo du α -quantile q_α :

$\hat{q}_{N,\alpha} = Y_{[N\alpha]}^*$, où $[\cdot]$ est la partie entière supérieure.

- Exemple : si $N = 100$ et $\alpha = 0.95 \rightarrow \hat{q}_{N,\alpha} = Y_{95}^*$.
si $N = 200$ et $\alpha = 0.95 \rightarrow \hat{q}_{N,\alpha} = Y_{190}^*$.

- Cette méthode n'a de sens que si $\frac{1}{N} < \alpha < 1 - \frac{1}{N}$

- Propriétés de $\hat{q}_{N,\alpha}$:

- Estimateur asymptotiquement sans biais: $\mathbb{E}[\hat{q}_{N,\alpha}] \xrightarrow{N \rightarrow \infty} q_\alpha$.
- Sa variance est d'autant plus grande que l'on cherche à évaluer un quantile extrême.
- $\sqrt{N} (\hat{q}_{N,\alpha} - q_\alpha) \xrightarrow{N \rightarrow \infty} \mathcal{N} \left(0, \frac{\alpha(1-\alpha)}{f_{y(x)}(q_\alpha)} \right)$.
 \rightarrow Malheureusement, $f_{y(x)}$ est inconnue donc cette propriété est inutile pour la construction d'un intervalle de confiance asymptotique.

- L'erreur d'estimation tendant vers une loi normale centrée,

$$\mathbb{P}\left(Y_{[N\alpha]}^* \geq q_\alpha\right) = 1/2, \quad \text{pour } N \text{ grand}$$

⇒ La probabilité de sous-estimation ou de sur-estimation du quantile est la même (problème pour des études de sûreté...)!

- Afin de proposer un estimateur plus conservatif, il a été proposé de prendre une **marge**, en prenant l'indice k plutôt que $[N\alpha]$, tel que:

$$\mathbb{P}(q_\alpha \leq Y_k^*) \geq \beta, \quad \beta > 1/2 \Leftrightarrow \mathbb{P}(\mathbb{P}(y(\mathbf{x}) \leq Y_k^*) \geq \alpha) \geq \beta.$$

→ $\alpha\%$ des valeurs possibles de $y(\mathbf{x})$ sont inférieures à Y_k^* avec un risque de se tromper de $1 - \beta$.

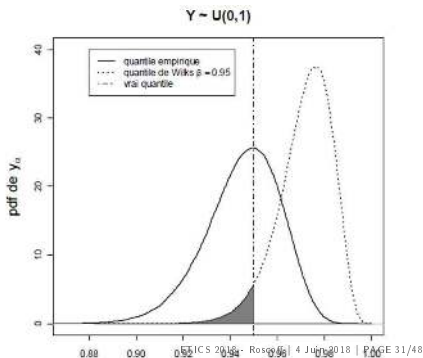
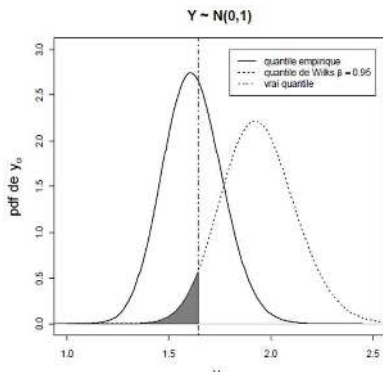
- On cherche la valeur k telle que $\mathbb{P}(q_\alpha \leq Y_k^*) \geq \beta$.
- Pour cela, on peut remarquer que:

$$\begin{aligned}\mathbb{P}(q_\alpha \leq Y_k^*) &= \mathbb{P}(\text{"au moins } k \text{ valeurs plus grandes que } q_\alpha \text{"}) \\ &= \sum_{n=0}^{k-1} \mathbb{P}(\text{"exactement } n \text{ valeurs plus petites que } q_\alpha \text{"}) \\ &= \sum_{n=0}^{k-1} C_N^n \alpha^n (1 - \alpha)^{N-n},\end{aligned}$$

si bien que k est le plus petit entier vérifiant:

$$\sum_{n=0}^{k-1} C_N^n \alpha^n (1 - \alpha)^{N-n} \geq \beta.$$

- On se donne un échantillon $\{Y_n\}_{n=1,\dots,N}$ de même loi de taille $N = 200$.
- Estimation de Y_k^* et $Y_{[N\alpha]}^*$.
- On réitère cela 10^6 fois.
- On peut tracer les pdf du quantile de Wilks et du quantile empirique.



- Pour α et β fixés, la méthode de Wilks permet de trouver la taille minimale N_{min} de l'échantillon nécessaire pour affirmer que :

$$\mathbb{P}(\mathbb{P}(y(\mathbf{x}) \leq Y_N^*) \geq \alpha) \geq \beta, \quad Y_N^* = \max_n(Y_n).$$

α	0.5	0.8	0.9	0.9	0.95	0.95	0.99	0.99
β	0.95	0.95	0.95	0.99	0.95	0.99	0.95	0.99
N_{min}	5	14	29	44	59	90	299	459

- Pour α , β et N fixés, la méthode de Wilks permet de trouver le rang du quantile (et donc sa valeur). Pour $\alpha = 0.95$ et $\beta = 0.95$:

N	50	59	80	100	130	170	220	300	400	500	800	1000
k	×	59	80	99	128	167	215	292	388	484	771	962
$[N\alpha]$	48	57	76	95	124	162	209	285	380	475	760	950

- Méthode robuste s'appliquant à **tout type de distribution** (Seule contrainte : l'échantillon $\{Y_n\}_{n=1,\dots,N}$ doit être aléatoire.)

- 1 Introduction
- 2 Garantir une probabilité de dépassement de seuil
- 3 Garantir un quantile
- 4 Sensibilité aux choix de modélisation
- 5 Conclusion

- La valeur de $P_f = \mathbb{P}(y(\mathbf{x}) \geq q)$ est complètement régie par le choix/l'identification des lois de \mathbf{x} et de $y(\mathbf{x})|\mathbf{x}$.
- A distributions fixées, il est possible d'introduire des estimateurs de P_f , puis de comparer le niveau de risque acceptable α à une transformation de ces estimateurs pour s'assurer d'un niveau de confiance β . Deux configurations peuvent alors se produire:
 - 1 P_f semble plus grand que α (ou du moins la confiance n'est pas suffisante sur le fait que P_f soit plus petit que α). On est dans une configuration "**négative**". **Sur quels paramètres se concentrer pour réduire P_f (sensibilité)?**
 - 2 P_f semble plus petit que α avec une confiance "raisonnable". On est dans une configuration "**positive**". Quitte à motiver de futures études sur ces lois, **sur quels paramètres se concentrer pour maximiser le crédit apporté à cette estimation de P_f (robustesse)?**

- La plupart des analyses de sensibilité/robustesse sont basées sur des évaluations du code **dédiées**, en des points bien définis.
- Dans ce cas, l'ensemble des calculs ne peut être utilisé, **à la fois**, pour l'évaluation de P_f et l'analyse de sa sensibilité/robustesse par rapport aux différents paramètres.
- Ceci explique que pour un grand nombre d'études de garantie, peu d'analyses de sensibilité/robustesse des résultats sont présentées.

⇒ L'idée de cette partie est de proposer quelques pistes pour ces études, en ne se basant que sur les calculs effectués pour l'estimation de P_f .

- Pour trouver les grandeurs ayant le plus d'influence sur P_f , il est intéressant de quantifier l'impact sur P_f de fixer x_i à une certaine valeur x_i^* .
- Plus $(P_f - \mathbb{P}(y(\mathbf{x}) > S \mid x_i = x_i^*))^2$ est grand, plus x_i a de chances de jouer un rôle sur P_f .
- En moyennant cette quantité sur x_i , la grandeur

$$\mathbb{E} \left[(\mathbb{P}(y(\mathbf{x}) > S) - \mathbb{P}(y(\mathbf{x}) > S \mid x_i)) \right]^2 = \mathbb{V}_{x_i} \left[\mathbb{E}_{\mathbf{x}_{-i}} [1_{y(\mathbf{x}) > S} \mid x_i] \right]$$

permet d'analyser la sensibilité de P_f à chaque entrée du modèle x_i .

- On retrouve ici les indices de Sobol d'ordre 1 de la fonction $1_{y(\mathbf{x}) > S}$

$$s_i := \frac{\mathbb{V}_{x_i} \left[\mathbb{E}_{\mathbf{x}_{-i}} [1_{y(\mathbf{x}) > S} \mid x_i] \right]}{\mathbb{V}_{\mathbf{x}} [1_{y(\mathbf{x}) > S}]}$$

- Par construction, s_i caractérise la part de variance de $1_{y(\mathbf{x}) > S}$ expliquée par x_i individuellement. La part de variance expliquée par x_i en intégrant les interactions avec les autres composantes de \mathbf{x} est donnée par l'indice de Sobol total t_i , qui vérifie:

$$t_i := 1 - \frac{\mathbb{V}_{\mathbf{x}_{-i}} [\mathbb{E}_{x_i} [1_{y(\mathbf{x}) > S} \mid \mathbf{x}_{-i}]]}{\mathbb{V}_{\mathbf{x}} [1_{y(\mathbf{x}) > S}]}$$

- En remarquant que:

$$s_i = \frac{P_f}{1 - P_f} \mathbb{V}_{x_i} \left[\frac{f_{x_i|y(\mathbf{x}) > S}(x_i)}{f_{x_i}(x_i)} \right], \quad t_i = 1 - \frac{P_f}{1 - P_f} \mathbb{V}_{\mathbf{x}_{-i}} \left[\frac{f_{\mathbf{x}_{-i}|y(\mathbf{x}) > S}(\mathbf{x}_{-i})}{f_{\mathbf{x}_{-i}}(\mathbf{x}_{-i})} \right],$$

et en estimant les PDFs $f_{x_i|y(\mathbf{x}) > S}$ et $f_{\mathbf{x}_{-i}|y(\mathbf{x}) > S}$ via des approches non-paramétriques à noyaux (\leftrightarrow "généralisation continue des histogrammes"), il est alors possible de quantifier de manière relativement précise le rôle de chaque paramètre d'entrée sur P_f .

Remarquons maintenant que pour toute fonction f_1, \dots, f_D positives définies sur \mathbb{R} et d'intégrale 1, la quantité $h(f_1, \dots, f_D)$ telle que:

$$h(f_1, \dots, f_D) := \int_{\mathbb{X}} 1_{y(\mathbf{x}) > S} \prod_{j=1}^D f_j(x_j) d\mathbf{x} \quad (1)$$

définit la probabilité que $y(\mathbf{x})$ soit plus grande que S , sous la condition que la PDF de \mathbf{x} est égale à $\prod_{j=1}^D f_j$.

Pour $0 \leq \delta < 1$, on note alors $\mathcal{F}_i(\delta)$ un ensemble de PDFs correspondant à des perturbations d'"amplitude δ " de f_{x_i} (par ex., modification de la moyenne et de la variance pour une variable gaussienne), et:

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

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

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

Par construction, p^δ correspond au pire cas en perturbant toutes les PDFs, p_i^δ en ne perturbant que celle de u_i , et p_{-i}^δ en perturbant toutes les PDFs des composantes de x sauf la $i^{\text{ème}}$.

L'influence de chaque entrée sur la différence $p^\delta - P_f$ peut ainsi être caractérisée par la comparaison des deux indices suivants:

$$\zeta_i^\delta := \frac{p_i^\delta - P_f}{p^\delta - P_f},$$

$$\tau_i^\delta := 1 - \frac{p_{-i}^\delta - P_f}{p^\delta - P_f}.$$

- ζ_i^δ caractérise le pourcentage d'augmentation dû aux effets individuels,
- τ_i^δ caractérise le pourcentage d'augmentation dû aux effets individuels et couplés.

La valeur de δ peut être choisie pour garantir une valeur de $p^\delta - P_f$ donnée (par exemple $p^\delta = 2 \times P_f$ dans la suite).

Il est possible d'estimer toutes ces quantités sans nouveaux appels au code en utilisant des méthodes basées sur le tirage d'importance

En première application, on considère la fonction polynômiale suivante:

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

et on s'intéresse à l'estimation de $P_f = \mathbb{P}(y(\mathbf{x}) > S)$, $S = 250$.

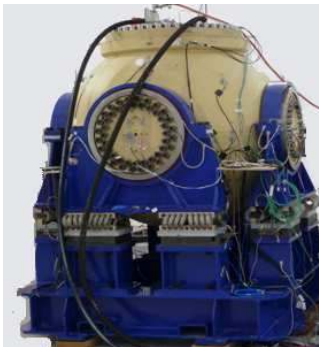
On trouve:

$$\mathbb{P}(\mathbb{P}(y(\mathbf{x}) > q) \in [0.0080; 0.0092]) \approx 95\%.$$

On cherche à identifier la sensibilité de cette estimation de P_f aux paramètres x_i

		$\hat{s}_i(\%)$	$\hat{t}_i(\%)$	$\hat{\zeta}_i^\delta(\%)$	$\hat{\tau}_i^\delta(\%)$
i=1	Référence	[1;21]	[94;107]	[57;61]	[70;73]
	Code+MP	[6;18]	[97;99]	[54;64]	[67;75]
	GP+MP	[8;15]	[96;99]	[53;64]	[67;76]
i=2	Référence	[-14;10]	[74;87]	[18;21]	[28;31]
	Code+MP	[0;1]	[41;81]	[16;22]	[25;32]
	GP+MP	[0;1]	[66;91]	[15;23]	[24;35]
i=3	Référence	[-9;8]	[47;56]	[7;9]	[12;15]
	Code+MP	[0;0]	[41;78]	[6;11]	[10;18]
	GP+MP	[0;0]	[50;83]	[6;11]	[10;18]

- $\delta = 0.05$ pour que $p^\delta \approx 2P_f$. La procédure est répétée 100 fois.
- "Référence" : MC basée sur 6.10^6 évaluations.
- "Code+MP" est basé sur environ 7200 code évaluations.
- "GP+MP" est basé sur **50 appels au code** (30 pour le DoE initial, 20 pour l'enrichissement).



Entrées

- x_1 : rayon interne de la cuve,
- x_2 : épaisseur,
- x_3 : incertitudes sur l'amplitude de $P(t)$,
- x_4 : incertitudes sur la chrono de $P(t)$,
- x_5 : module de Young de la cuve,
- x_6 : limite d'élasticité de la cuve,
- x_7 : module de Young du bouchon,
- x_8 : limite d'élasticité du bouchon.

Sortie

$y \rightarrow$ déformation plastique maximale en temps et en espace.

$$P_f = \mathbb{P}(y(\mathbf{x}) > 0.05) \approx 0.0015.$$

	$\hat{s}_i(\%)$	$\hat{t}_i(\%)$	$\hat{\zeta}_i^\delta(\%)$	$\hat{\tau}_i^\delta(\%)$
$i=1$	[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: "GP+MP" est basé sur **1286** appels au code (500 pour le DoE initial et 786 pour l'enrichissement itératif). Les valeurs entre crochets sont des intervalles à 95%, intégrant l'incertitude due au métamodèle et au caractère approché de la probabilité, mais en se basant sur le même métamodèle. $\delta = 0.077$ si bien que $p^\delta \approx 2P_f$.

⇒ le seuil est dépassé pour des configurations nécessitant des combinaisons bien particulières de tous les paramètres.

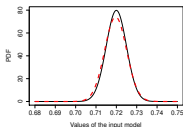
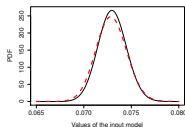
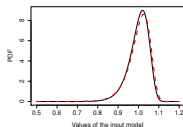
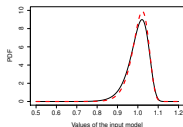
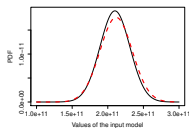
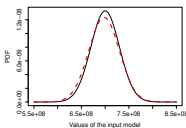
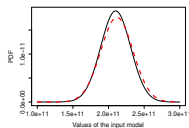
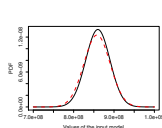
(a) z_1 (b) z_2 (c) z_3 (d) z_4 (e) z_5 (f) z_6 (g) z_7 (h) z_8

Figure: Comparaison entre les PDFs initiales (en noir) et les PDFs perturbées (en rouge pointillé) conduisant à la multiplication de P_f par un facteur 2.

- 1 Introduction
- 2 Garantir une probabilité de dépassement de seuil
- 3 Garantir un quantile
- 4 Sensibilité aux choix de modélisation
- 5 Conclusion

- Garantir par la simulation, c'est chercher à évaluer une double probabilité: la première associée à un risque, la seconde pour attribuer une confiance à l'estimateur.
- Le problème de garantie peut s'exprimer sur le quantile aussi bien que sur la probabilité de dépasser un seuil.
- Plus le risque est faible, et plus la garantie demande un nombre d'appels au code conséquent (\Rightarrow on passe souvent par des métamodèles pour les faibles probabilités, dont l'incertitude doit aussi être propagée dans le modèle).
- Il est difficile de garantir une quantité observable, et encore plus difficile de garantir une quantité non observable.

- La garantie s'effectue bien sous conditions:
 - que l'on ait confiance en la théorie des probabilités pour modéliser l'incertain,
 - que l'on sache générer des nombres aléatoires,
 - que les sources d'incertitudes aient été bien identifiées et modélisées.
- Afin d'être davantage conservatif vis-à-vis d'incertitudes sur les lois des entrées, on notera le développement actuel des approches dites "robust Bayes", qui cherchent à résoudre des problèmes du type:

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

où \mathbb{F}_y et \mathbb{F}_x sont des ensembles de lois sous contrainte de la "seule vraie" connaissance sur \mathbf{x} et $y(\mathbf{x}) \mid \mathbf{x}$...

- La garantie s'effectue bien sous conditions:
 - que l'on ait confiance en la théorie des probabilités pour modéliser l'incertain,
 - que l'on sache générer des nombres aléatoires,
 - que les sources d'incertitudes aient été bien identifiées et modélisées.
- Afin d'être davantage conservatif vis-à-vis d'incertitudes sur les lois des entrées, on notera le développement actuel des approches dites "robust Bayes", qui cherchent à résoudre des problèmes du type:

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

où \mathbb{F}_y et \mathbb{F}_x sont des ensembles de lois sous contrainte de la "seule vraie" connaissance sur \mathbf{x} et $y(\mathbf{x}) \mid \mathbf{x}$...

Merci pour votre attention!

