

# Analyse de risque

**G. Perrin**

guillaume.perrin@univ-eiffel.fr

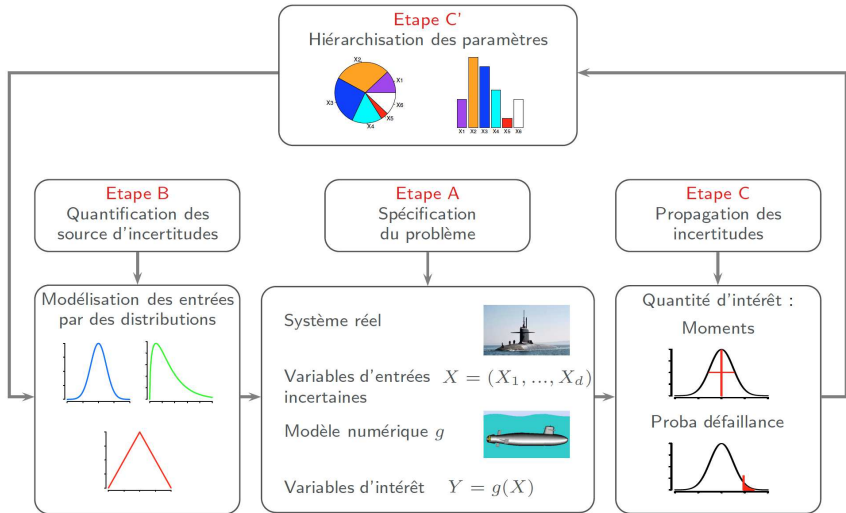
Année 2022-2023



# Plan de la séance

- 1 Introduction
- 2 Echantillonnage conditionné
- 3 Quelques méthodes approchées
  - Méthodes FORM/SORM
  - Méthodes directionnelles

# Introduction - Schéma général



# Introduction - formulation probabiliste

La fiabilité d'un système (physique, chimique, mécanique, financier) fait référence à plusieurs grandeurs :

- $\mathbf{x} = (x_1, \dots, x_d) \leftrightarrow$  paramétrage du système, que l'on modélise comme un **vecteur aléatoire**, dont la PDF est notée  $f_{\mathbf{x}}$ ,
- $y \leftrightarrow$  quantité d'intérêt pour la surveillance du bon fonctionnement du système considéré,
- $S \leftrightarrow$  seuil à ne pas dépasser.

Pour garantir le bon fonctionnement du système, il convient alors de calculer la probabilité de défaillance, notée  $p_f$ , et vérifiant :

$$p_f := \mathbb{P}(y(\mathbf{x}) > S),$$

puis de décider si cette valeur est admissible ou non (vis-à-vis de normes de sécurité par exemple).

# Introduction - plusieurs seuils

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

- Le système est suffisamment "sûr" si  $p_f \leq \alpha$ .
- Si  $\hat{p}_n$  est un **estimateur** statistique de  $p_f$  basé sur  $n$  évaluations de  $y$ , on peut dire que le système est considéré comme suffisamment "sûr" si  $\hat{p}_n + c(n, \alpha, \beta) \leq \alpha$ , où  $c(n, \alpha, \beta)$  est choisie de manière à éviter toute fausse certification avec probabilité  $1 - \beta$  :

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

La garantie du système repose alors sur plusieurs constantes

- $S = 0 \leftrightarrow$  seuil sur  $y$  à ne pas dépasser,
- $\alpha \leftrightarrow$  risque "acceptable" (en référence à  $S$ ),
- $\beta \leftrightarrow$  niveau de confiance (remplacement de  $p_f$  par  $\hat{p}_n$ ),
- $c(n, \alpha, \beta) \leftrightarrow$  marge de sécurité (en référence à  $\hat{p}_n, S, \alpha, \beta$ ).

# Introduction - le cas Monte Carlo

- Supposons alors que  $\mathbf{x}_1, \dots, \mathbf{x}_n$  soient  $n$  réalisations i.i.d. de  $\mathbf{x}$ ,
- $\hat{p}_n := \frac{1}{n} \sum_{i=1}^n 1_{y(\mathbf{x}_i) < 0}$  est l'estimateur classique MC de  $p_f$ .
- En remarquant que  $\sqrt{n}(\hat{p}_n - p_f) \xrightarrow{\mathcal{L}} \mathcal{N}(0, p_f(1 - p_f))$ , on trouve que

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

est un choix naturel pour retrouver (aymptotiquement) le critère de certification voulu, avec  $\phi_{1-\beta}$  le quantile de niveau  $(1 - \beta)$  d'un v.a. gaussienne :

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

Exercice : montrer le résultat précédent.

# Introduction - le cas Monte Carlo

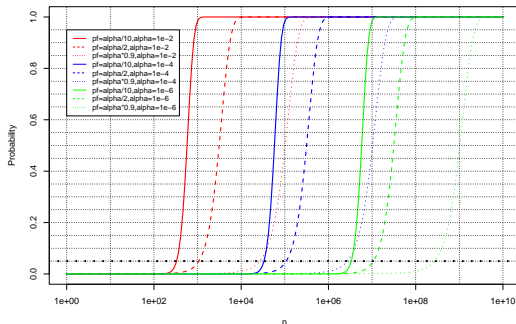
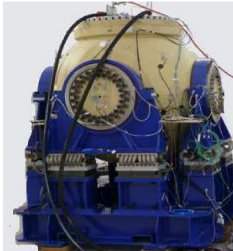


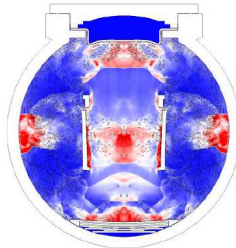
FIGURE:  $n \mapsto \mathbb{P}(\hat{p}_n + c(n, \alpha, \beta = 5\%) \leq \alpha)$

⇒ De très nombreuses évaluations de  $y$  sont nécessaires pour obtenir une confiance raisonnables sur les résultats ( $n \approx 10\alpha$  quand  $p_f \approx 0.1\alpha$ ,  $n \approx 50\alpha$  quand  $p_f \approx 0.5\alpha$ ,  $n \approx 10^3\alpha$  quand  $p_f \approx 0.9\alpha$ ).

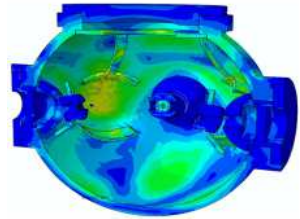
# Un exemple



(a) Conteneur expérimental



(b) Hydrodynamique



(c) Dynamique des structures

**FIGURE:** Conteneur pressurisé pour des expériences de dynamique rapide

## Problématique industrielle

Comment s'assurer que la valeur maximale en temps et espace de l'endommagement plastique restera en dessous d'une valeur critique ?



# Introduction - remarques finales

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  $\mathbf{x}$  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  $y(\mathbf{x})$  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.

Généralement, on parle de **fiabilité** ou d'évènement **rare** lorsque la probabilité à évaluer,  $p_f$ , est de l'ordre de / inférieure à  $10^{-3}$ .

# Introduction - remarques finales

- On appelle généralement "queue de distribution" la zone de valeurs de  $y$  d'intérêt, c'est à dire les valeurs autour de / supérieures à  $S$ .
- Dans ce cours, on insiste sur le fait que la PDF  $f_y$  n'est accessible que de manière indirecte, c'est à dire qu'à travers celle de  $\mathbf{x}$  :

$$f_y(u)du := \int_{\mathbb{R}} 1_{u \leq y \leq u+du} f_y(y) dy = \int_{\mathbb{R}^d} 1_{u \leq y(\mathbf{x}) \leq u+du} f_{\mathbf{x}}(\mathbf{x}) d\mathbf{x}.$$

## Problématique scientifique

L'idée est 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  $p_f$ .

# Plan de la séance

- 1 Introduction
- 2 Echantillonnage conditionné
- 3 Quelques méthodes approchées
  - Méthodes FORM/SORM
  - Méthodes directionnelles

# Echantillonnage conditionné - principe

Afin de réduire le nombre important d'évaluations requis par MC, 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  $p_f$ .

Exercice : calculer la moyenne, la variance et le coefficient de variation de ce nouvel estimateur, en supposant que l'on a effectué  $N_m$  tirages indépendants pour l'estimation de  $P_m$ .

# Echantillonnage conditionné - stratégie optimale

- 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".

# Echantillonnage conditionné - stratégies approchées

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

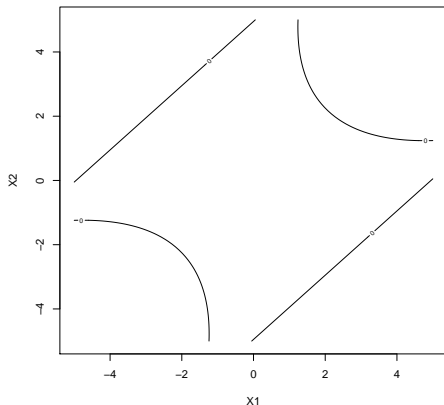
- les approches dites **"subset simulation"** : on choisit une valeur de  $p_0$  (souvent 0.1) et de  $N_0$  (entre 100 et 1000), puis on rééchantillonne itérativement les  $(1 - p_0)N_0$  plus faibles valeurs pour être plus grand que la  $p_0 N_0$  plus haute valeur. On s'arrête quand cette  $p_0 N_0$  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 :

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

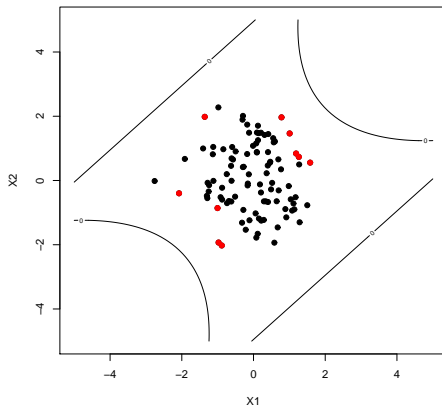
# Echantillonnage conditionné - Subset simulation

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



# Echantillonnage conditionné - Subset simulation

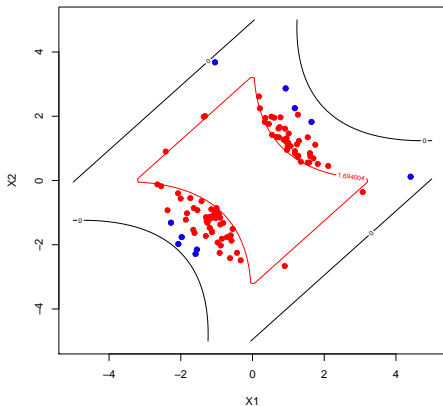
$p_0 = 0.1$ ,  $N_0 = 100$ . Rouge :  $p_0 N_0 = 10$  points  $> S_1$ .





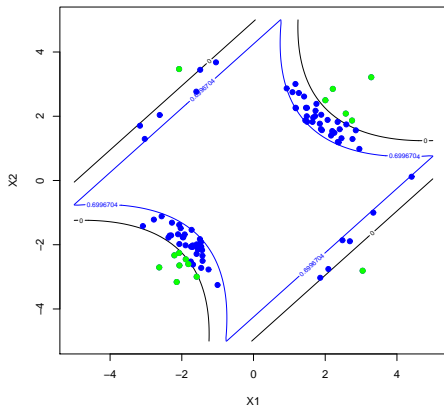
# Echantillonnage conditionné - Subset simulation

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



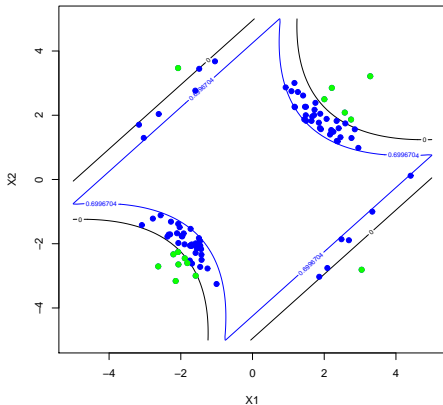
# Echantillonnage conditionné - Subset simulation

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



# Echantillonnage conditionné - Subset simulation

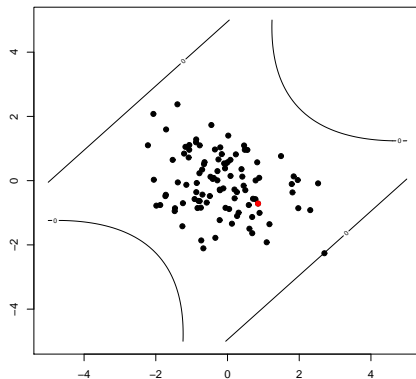
$p_0 = 0.1$ ,  $N_0 = 100$ . Vert : 15 points au dessus de  $S$ .



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

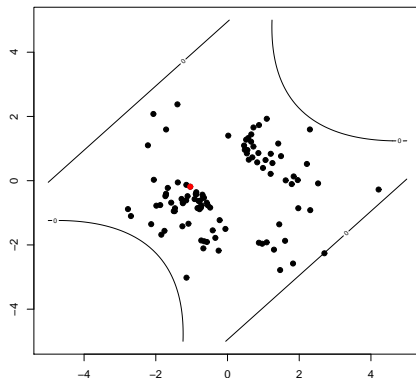
# Echantillonnage conditionné - Moving particle

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



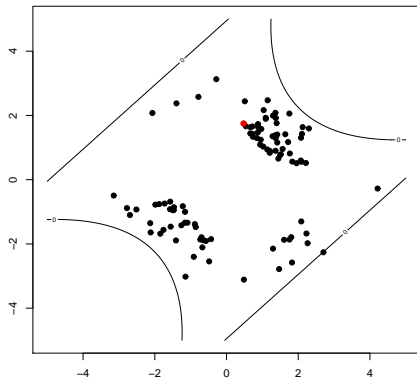
# Echantillonnage conditionné - Moving particle

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



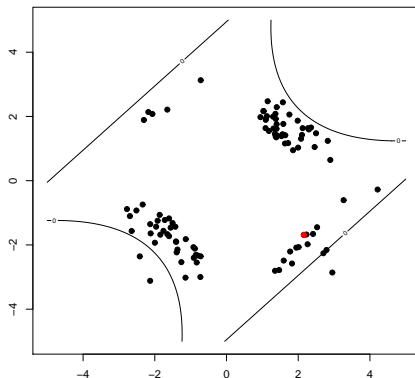
# Echantillonnage conditionné - Moving particle

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



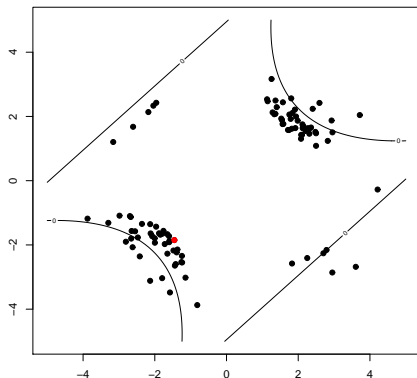
# Echantillonnage conditionné - Moving particle

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



# Echantillonnage conditionné - Moving particle

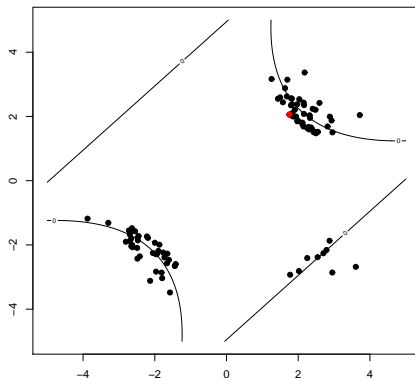
$N_0 = 100$ . Rouge : point minimum après 400 déplacements.





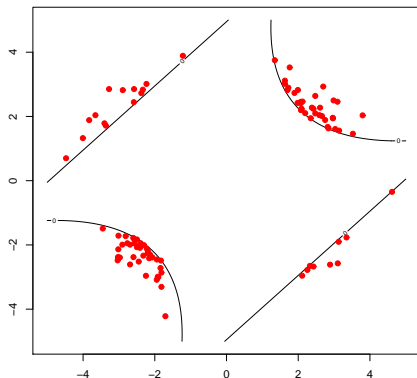
# Echantillonnage conditionné - Moving particle

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



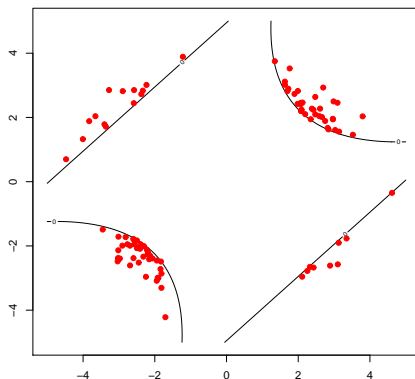
# Echantillonnage conditionné - Moving particle

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



# Echantillonnage conditionné - Moving particle

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



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

# Echantillonnage conditionné - avantages et désavantages

## Echantillonnage par Monte Carlo

- (+) facile à mettre en oeuvre.
- (-) demande un très grand nombre d'évaluations de  $y$ .

## Echantillonnage conditionné

- (+) demande un nombre a priori réduit d'évaluations de  $y$
- (-) plus difficile à mettre en oeuvre (nécessité de pouvoir tirer  $\mathbf{x}$  conditionnellement à un seuil donné sur  $y(\mathbf{x})$  - voir les méthodes MCMC dédiées).

# Echantillonnage conditionné - retour à la garantie

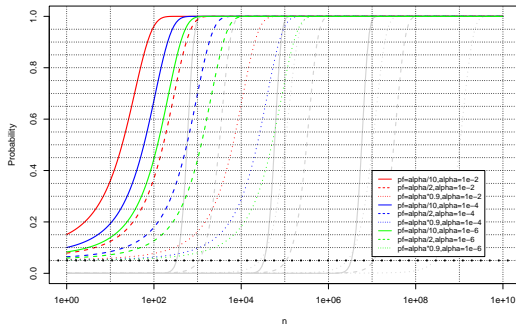


FIGURE:  $n \mapsto \mathbb{P}(\tilde{p}_N + c(Q, \alpha, \beta = 5\%) \leq \alpha)$ ,  $n = Q(1 - \log(p_f))$ .

⇒ On a besoin de beaucoup moins d'appels à  $y$  pour obtenir la même confiance sur les résultats dans le cas du Moving Particle (avec génération de points par  $Q = 10$  itérations MCMC), en particulier pour les faibles valeurs de  $p_f$ .

# Plan de la séance

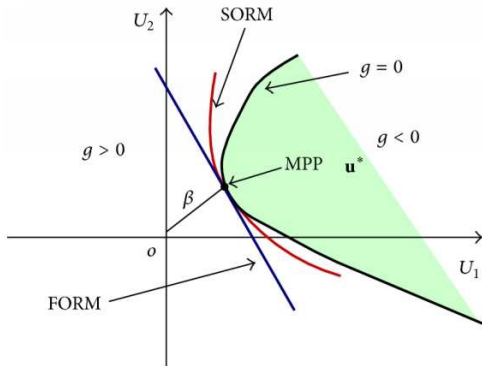
- 1 Introduction
- 2 Echantillonnage conditionné
- 3 Quelques méthodes approchées
  - Méthodes FORM/SORM
  - Méthodes directionnelles

# Méthodes approchées

Les méthodes approchées sont généralement basées sur :

- une transformation de l'espace des entrées dans l'espace gaussien centré normé :  $\mathbf{x} \rightarrow \mathbf{U}(\mathbf{x})$ , où  $\mathbf{U}$  est un v.a. centré gaussien de covariance la matrice identité (**transformation isoprobabiliste**),
- la recherche du (des) point(s) défaillant(s) le(s) plus "proche(s)" de l'origine dans l'espace transformé (**point(s) critique(s)**),
- l'approximation de la surface de défaillance par un hyper-plan (FORM), un parabolöide (SORM), une hypersphère (méthodes directionnelles),
- la déduction de la probabilité de défaillance à partir des propriétés d'invariance par rotation de la distribution gaussienne centrée réduite.

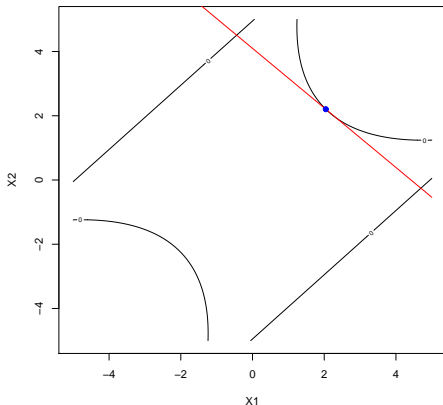
# Méthodes FORM/SORM - illustration graphique



$$\text{FORM : } p_f \approx \int_{\beta}^{+\infty} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) dx = \Phi(-\beta).$$



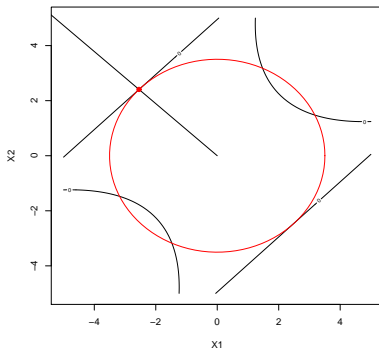
# Méthodes FORM/SORM - illustration graphique



$p_f \approx 0.0013$ . Commenter les avantages et désavantages d'une telle approche pour la fonction Waarts.

# Méthodes directionnelles - illustration graphique

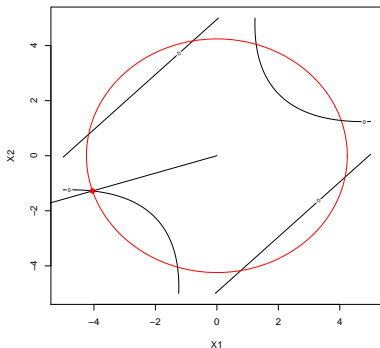
Point rouge : point le plus proche selon direction noire.



$$P_1 \approx 0.0021$$

# Méthodes directionnelles - illustration graphique

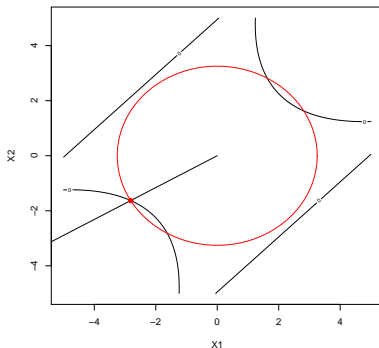
Point rouge : point le plus proche selon direction noire.



$$P_2 \approx 0.00012$$

# Méthodes directionnelles - illustration graphique

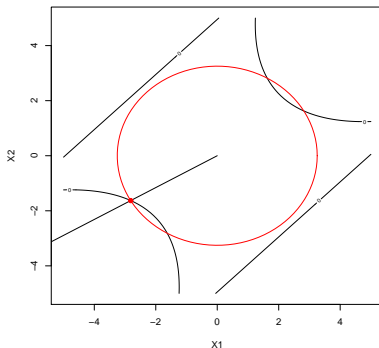
Point rouge : point le plus proche selon direction noire.



$$P_3 \approx 0.0050$$

# Méthodes directionnelles - illustration graphique

Point rouge : point le plus proche selon direction noire.



$$p_f \approx \frac{1}{N_{\text{iter}}} \sum_{i=1}^{N_{\text{iter}}} P_i \approx 0.0020 \text{ (pour } N_{\text{iter}} = 10).$$

$$P_3 \approx 0.0050$$

# Méthodes approchées

- (+) facile à mettre en oeuvre.
- (+) le problème d'intégration est remplacé par un problème de minimisation, nécessitant souvent beaucoup moins d'évaluations.
- (-) Aucune véritable garantie sur le résultat (difficile d'évaluer une confiance).

# Plan de la séance

- 1 Introduction
- 2 Echantillonnage conditionné
- 3 Quelques méthodes approchées
  - Méthodes FORM/SORM
  - Méthodes directionnelles