

DE LA RECHERCHE À L'INDUSTRIE



SIMULATION D'ÉVÉNEMENTS RARES

Application à l'estimation de probabilités et
de quantiles extrêmes

Forum incertitudes CEA/DAM | Clément WALTER

1 octobre 2014

■ Modélisation du problème :

- \mathbf{X} vecteur aléatoire des paramètres d'entrée du modèle, de loi connue
- g une fonction *boîte-noire* représentant le code, $g : \mathbb{R}^d \rightarrow \mathbb{R}$
- $Y = g(\mathbf{X})$ la variable aléatoire qualifiant l'état du système, de loi inconnue

■ Deux objectifs :

- Quantification d'incertitude : mesurer $F = \{\mathbf{x} \in \mathbb{R}^d \mid g(\mathbf{x}) > q\}$ au sens de la probabilité : $p = \mu^X(F) = \int_F d\mu^X = \int_{\mathbb{R}^d} \mathbb{1}_F(\mathbf{x}) d\mu^X$
- Norme de sécurité : retrouver q pour une probabilité de défaillance donnée

Problème

- $p = \mu^X(F) = \mu^Y([q; +\infty[) \ll 1$
- nécessité de g pour accéder à F ou μ^Y qui est une fonction coûteuse en temps de calcul

- 1 Aperçu des différentes stratégies
 - Méthodes de Monte-Carlo
 - Méthodes avec méta-modèle
- 2 Simulation d'évènements rares
 - Processus ponctuel
 - Estimation de probabilité
 - Estimation de quantile
- 3 Cas pratique
- 4 Conclusion

- 1 Aperçu des différentes stratégies
 - Méthodes de Monte-Carlo
 - Méthodes avec méta-modèle
- 2 Simulation d'évènements rares
 - Processus ponctuel
 - Estimation de probabilité
 - Estimation de quantile
- 3 Cas pratique
- 4 Conclusion

Échantillonner iid et compter le nombre de points : $p = \mathbb{E} [\mathbb{1}_{X > x_0}]$

Loi des grands nombres + Théorème Central Limite : $\delta^2 = \frac{1-p}{Np} \approx \frac{1}{Np}$

Exemple : estimer

$$p = \mathbb{P} [X > 2] \approx 2,28 \cdot 10^{-2}$$

avec $X \sim \mathcal{N}(0, 1)$

et $N = 10\,000$

$$\delta = 6,55\%$$

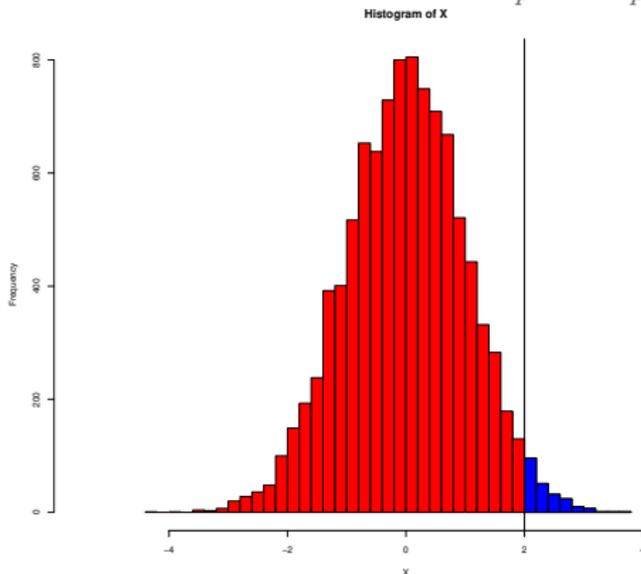
Code R :

```
X = rnorm(10000)
```

```
p_hat = mean(X > 2)
```

$$\hat{p} = 2,23 \cdot 10^{-2}$$

$$\hat{\delta} = 6,62\%$$



Information sur l'erreur | $N \geq 10^2/p$
Parallélisation

Approximer la fonction g par un méta-modèle \tilde{g} [2, 6, 7]

- Régression linéaire
- Processus Gaussien
- Polynôme de chaos...

Utiliser le métamodèle pour définir un classificateur \tilde{g}

Appliquer une méthode de Monte-Carlo sur $\tilde{F} = \{\mathbf{x} \mid \tilde{g}(\mathbf{x}) > q\}$

Apprentissage avec peu d'appels
 Estimation avec $N \geq 10^2/p$ possible
 erreur contrôlée

Qualité de l'approximation
 erreur incontrôlée

- 1 Aperçu des différentes stratégies
 - Méthodes de Monte-Carlo
 - Méthodes avec méta-modèle
- 2 Simulation d'évènements rares
 - Processus ponctuel
 - Estimation de probabilité
 - Estimation de quantile
- 3 Cas pratique
- 4 Conclusion

Idée

Travailler sur la variable aléatoire Y dont on ne connaît rien
On s'intéresse à des évènements extrêmes

Définition [10]

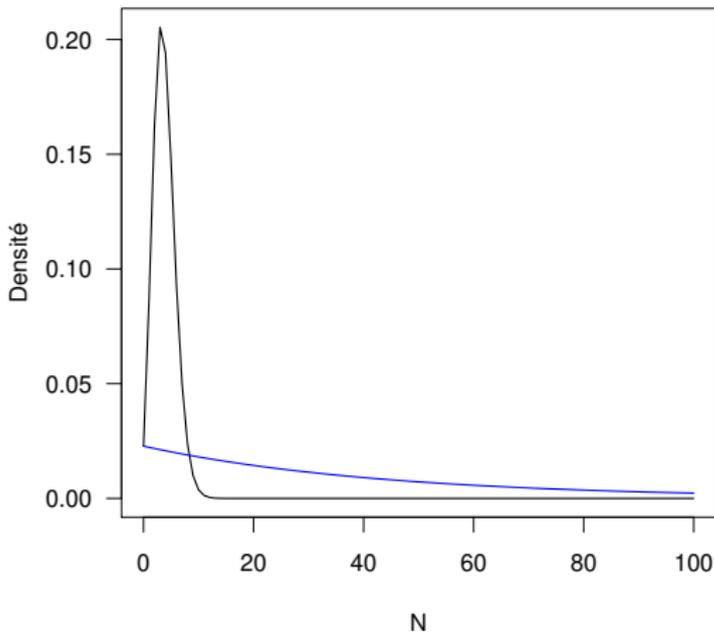
Soit Y une variable aléatoire réelle de loi μ^Y de fonction de répartition F_Y continue. On définit la chaîne de Markov :

$$Y_0 = -\infty$$

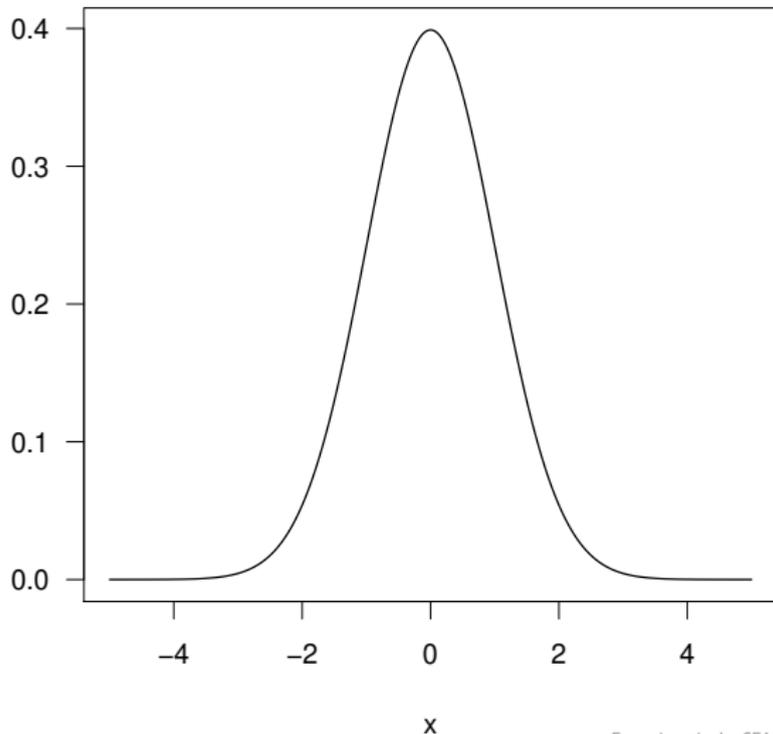
$$Y_{m+1} \sim \mu^Y(\cdot \mid Y > Y_m), \forall m \in \mathbb{N}$$

On peut montrer que le processus $T_m = -\log(P(Y > Y_m))$ est un **processus de Poisson de paramètre 1**

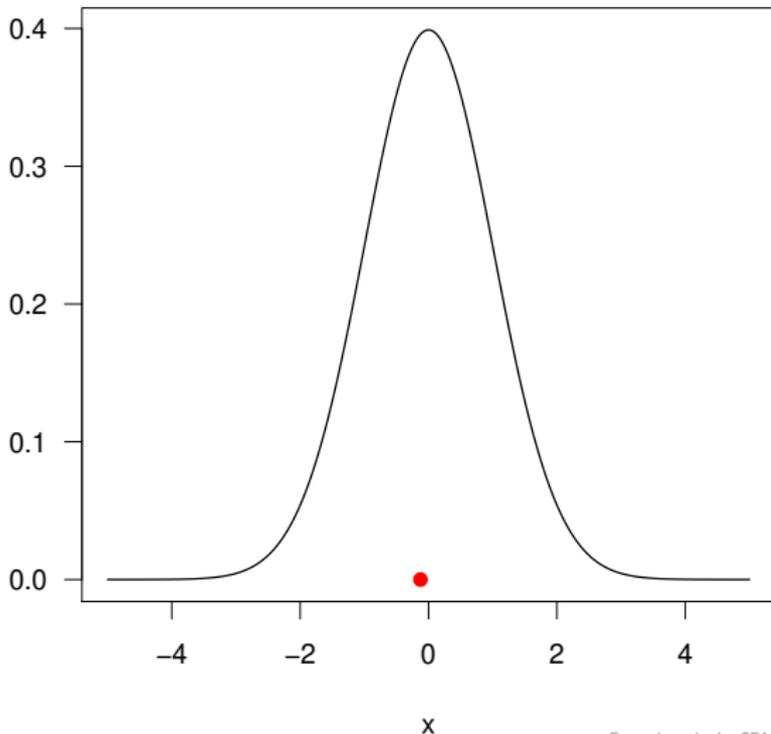
Comparaison des densités pour l'obtention d'un point au delà d'un seuil par tirage aléatoire *classique* (**loi géométrique en bleu**) et par cette chaîne de Markov (**loi de Poisson en noir**) pour $p = 2,28 \cdot 10^{-2}$



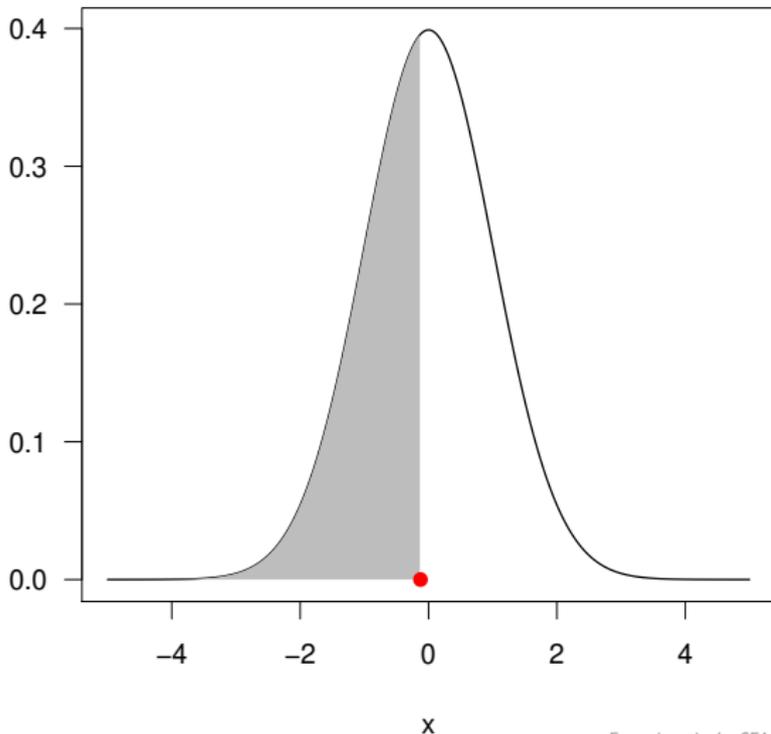
$Y \sim \mathcal{N}(0, 1)$; $p = P[Y > 2] \approx 2,28 \cdot 10^{-2}$; $1/p \approx 43,96$; $-\log p \approx 3,78$



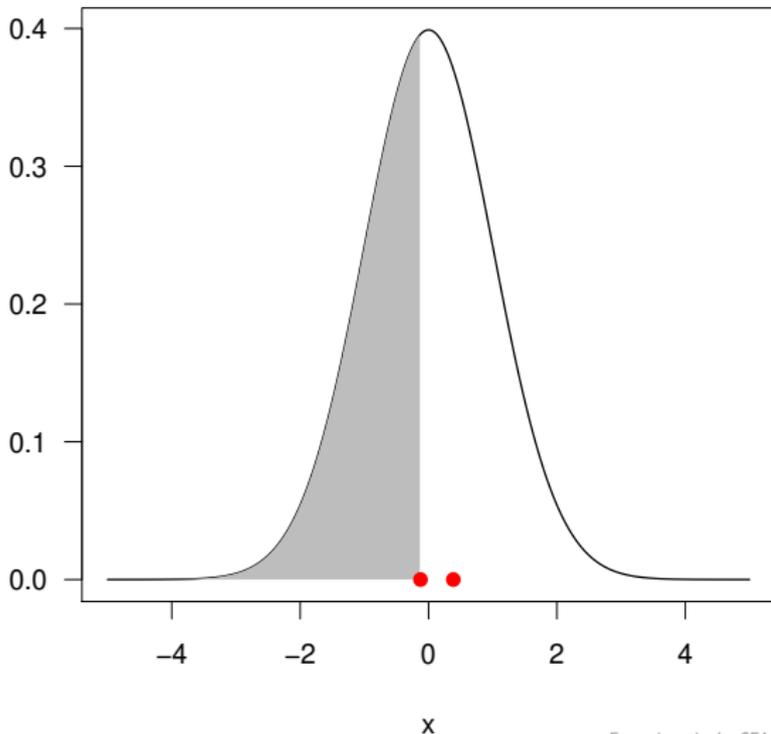
$Y \sim \mathcal{N}(0, 1)$; $p = P[Y > 2] \approx 2,28 \cdot 10^{-2}$; $1/p \approx 43,96$; $-\log p \approx 3,78$



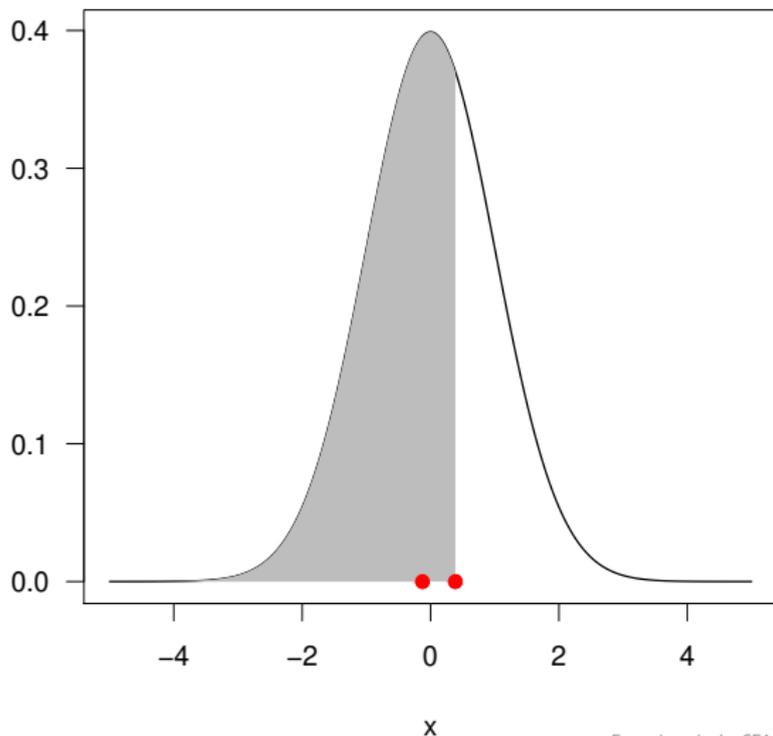
$Y \sim \mathcal{N}(0, 1)$; $p = P[Y > 2] \approx 2,28 \cdot 10^{-2}$; $1/p \approx 43,96$; $-\log p \approx 3,78$



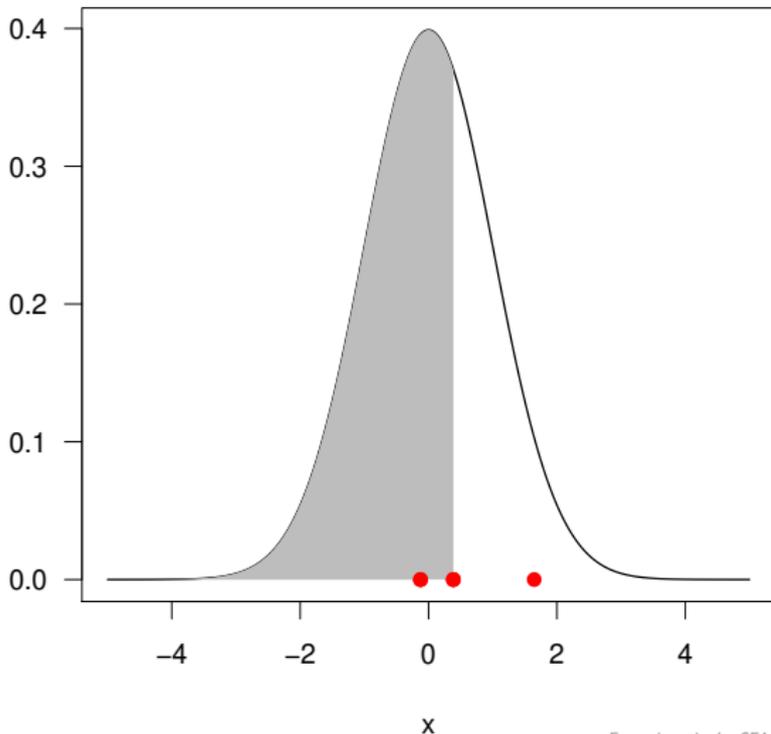
$Y \sim \mathcal{N}(0, 1)$; $p = P[Y > 2] \approx 2,28 \cdot 10^{-2}$; $1/p \approx 43,96$; $-\log p \approx 3,78$



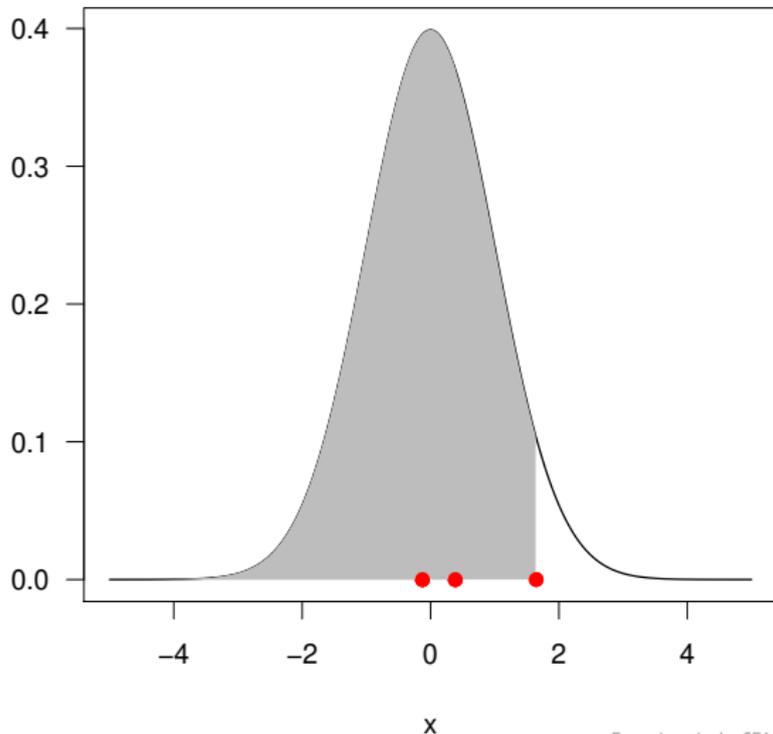
$Y \sim \mathcal{N}(0, 1)$; $p = P[Y > 2] \approx 2,28 \cdot 10^{-2}$; $1/p \approx 43,96$; $-\log p \approx 3,78$



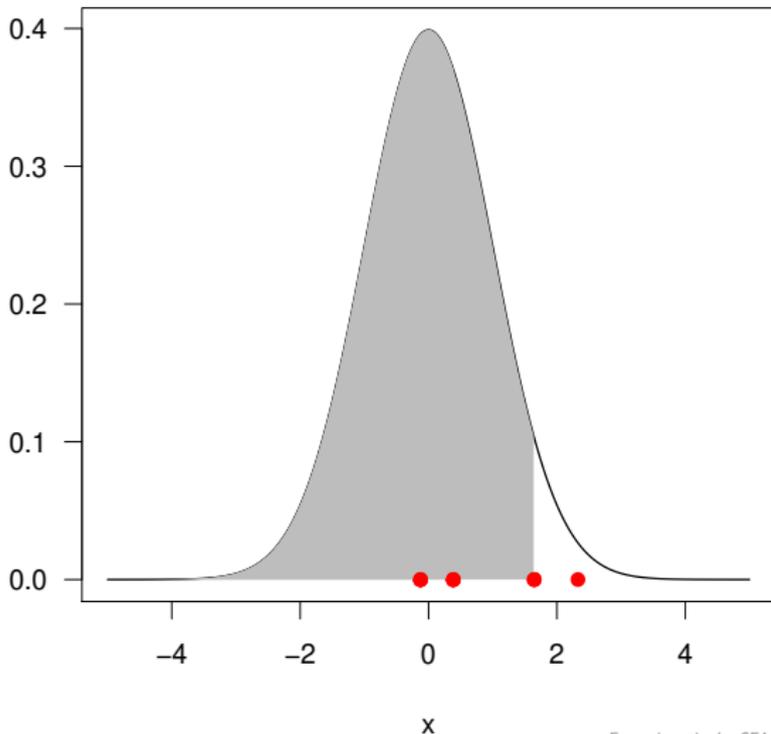
$Y \sim \mathcal{N}(0, 1)$; $p = P[Y > 2] \approx 2,28 \cdot 10^{-2}$; $1/p \approx 43,96$; $-\log p \approx 3,78$



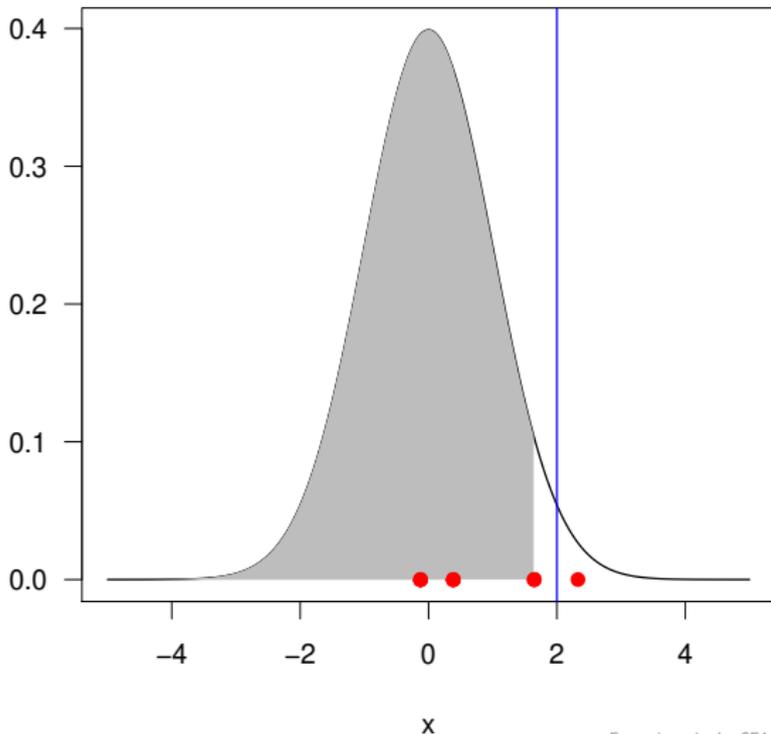
$Y \sim \mathcal{N}(0, 1)$; $p = P[Y > 2] \approx 2,28 \cdot 10^{-2}$; $1/p \approx 43,96$; $-\log p \approx 3,78$



$Y \sim \mathcal{N}(0, 1)$; $p = P[Y > 2] \approx 2,28 \cdot 10^{-2}$; $1/p \approx 43,96$; $-\log p \approx 3,78$



$Y \sim \mathcal{N}(0, 1)$; $p = P[Y > 2] \approx 2,28 \cdot 10^{-2}$; $1/p \approx 43,96$; $-\log p \approx 3,78$



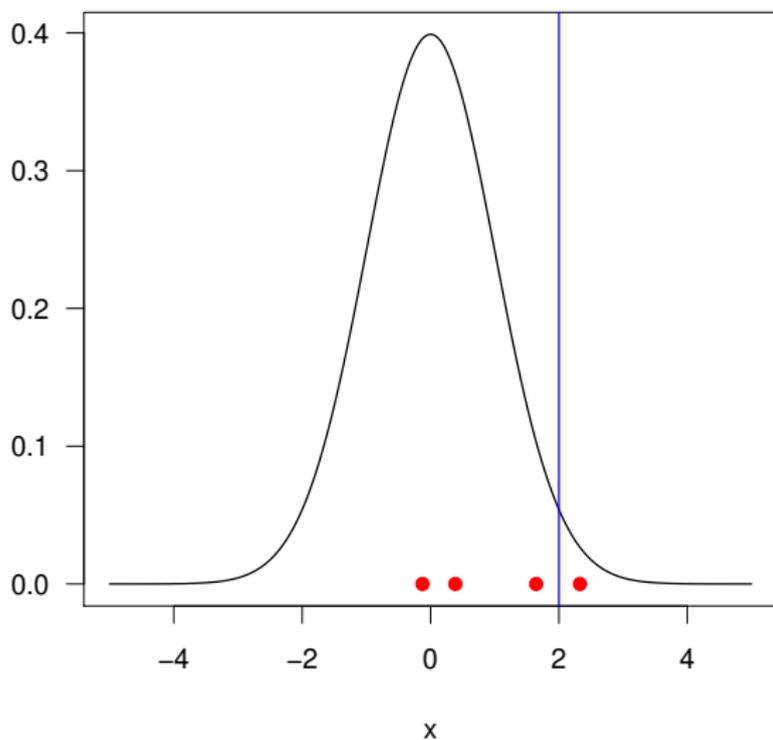
Concept [10]

Le nombre d'évènements avant le temps $t = -\log(P(Y > q))$ suit une loi de Poisson de paramètre t

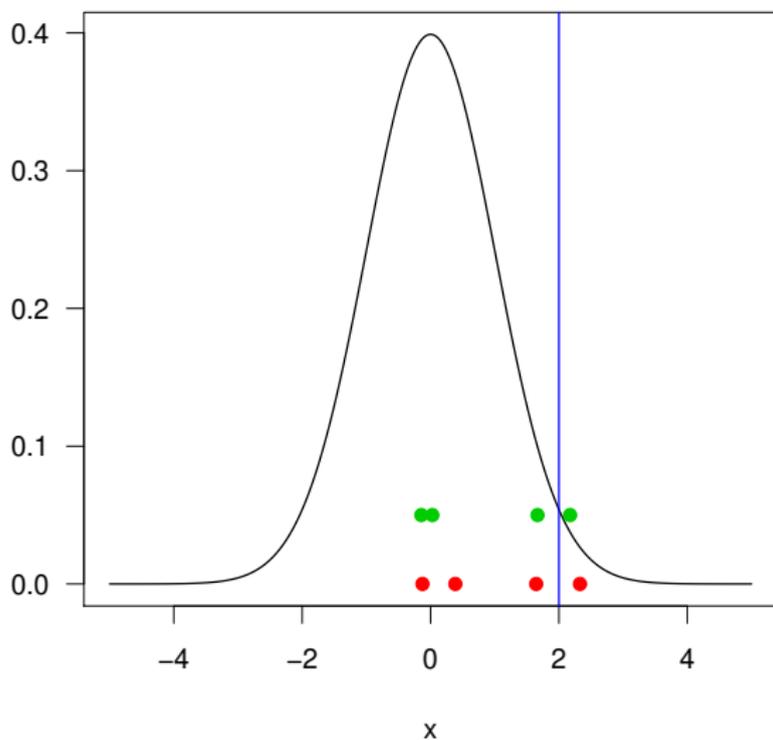
Simuler N processus et estimer ce paramètre

- Soit $(M_i)_{i=1..N}$ les N VA du nombre d'évènements au temps $t = -\log(p)$: $M_i \sim \mathcal{P}(-\log p)$; $M_q = \sum_{i=1}^N M_i \sim \mathcal{P}(-N \log p)$
- $-\log p \approx \frac{1}{N} \sum_{i=1}^N M_i = \frac{M_q}{N} \longrightarrow \hat{p} = \left(1 - \frac{1}{N}\right)^{M_q}$
- $E[\hat{p}] = p$; $\delta^2 = p^{-1/N} - 1 \approx \frac{\log 1/p}{N}$

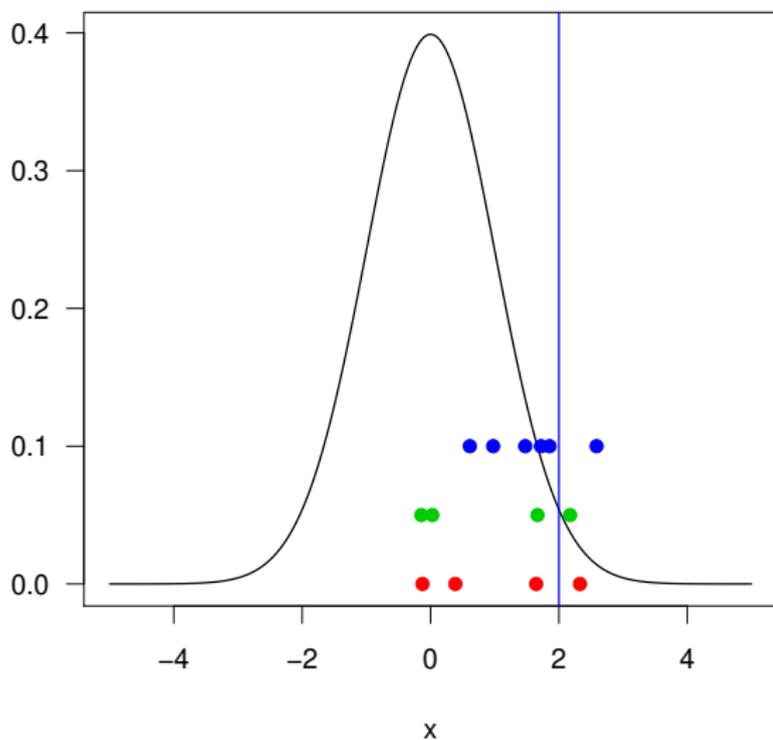
$$\hat{p} = (1 - 1/N)^{\{\text{nombre total de déplacements}\}}$$



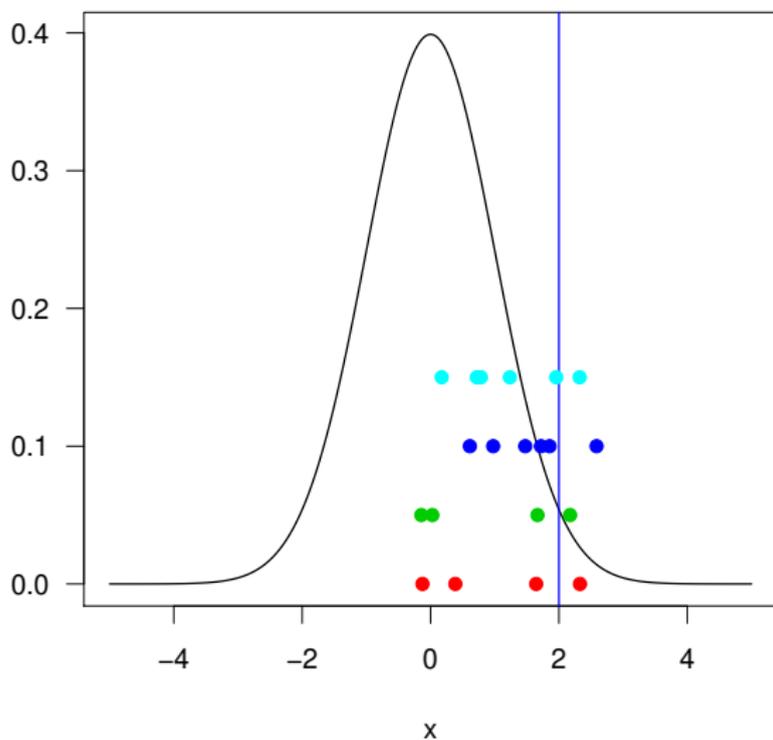
$$\hat{p} = (1 - 1/N) \{\text{nombre total de déplacements}\}$$



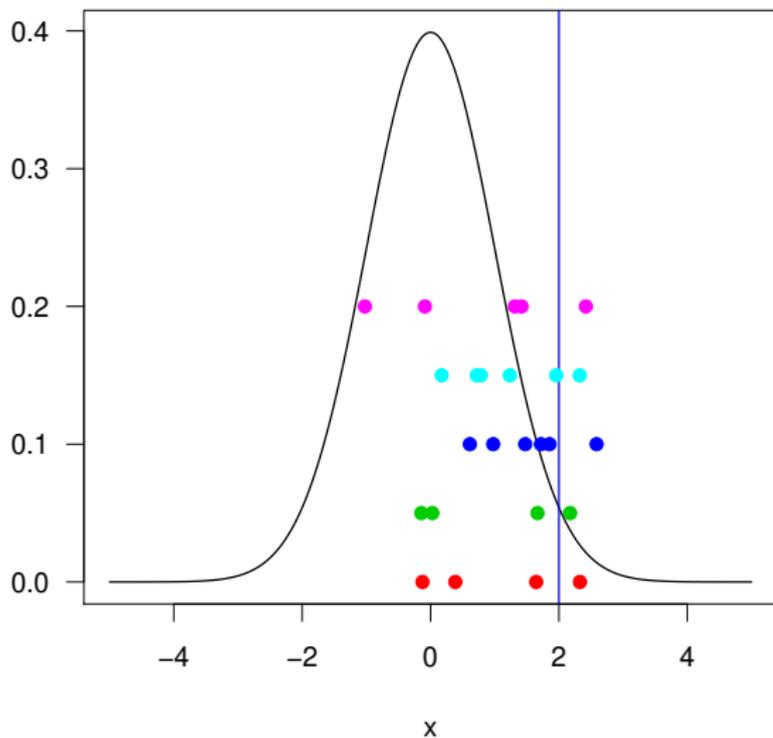
$$\hat{p} = (1 - 1/N) \{\text{nombre total de déplacements}\}$$



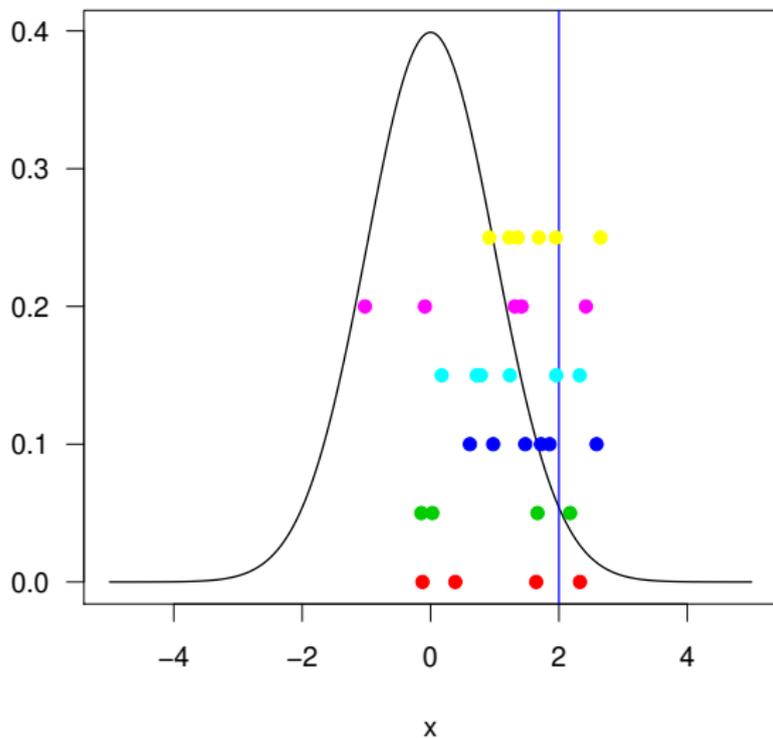
$$\hat{p} = (1 - 1/N) \{\text{nombre total de déplacements}\}$$



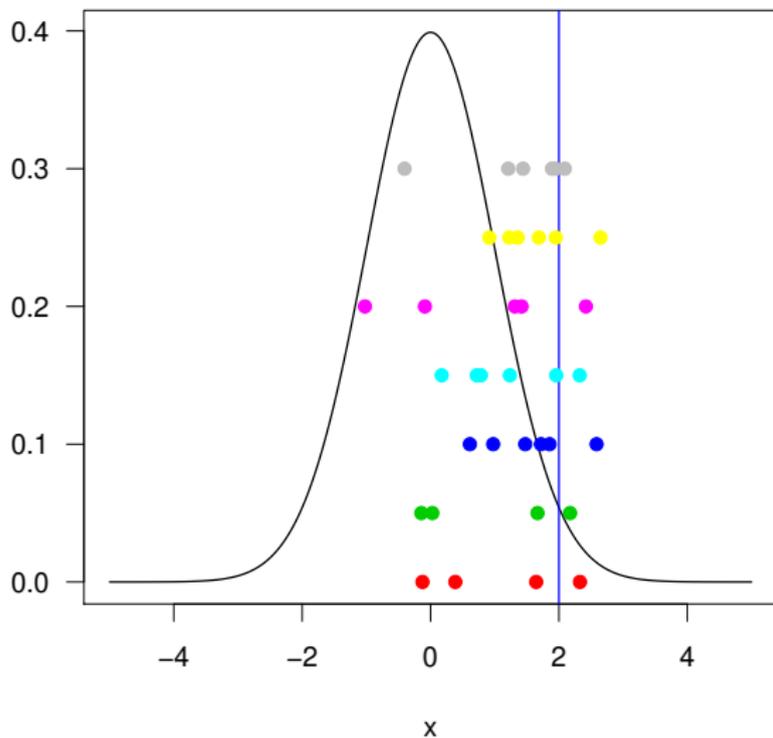
$$\hat{p} = (1 - 1/N) \{\text{nombre total de déplacements}\}$$



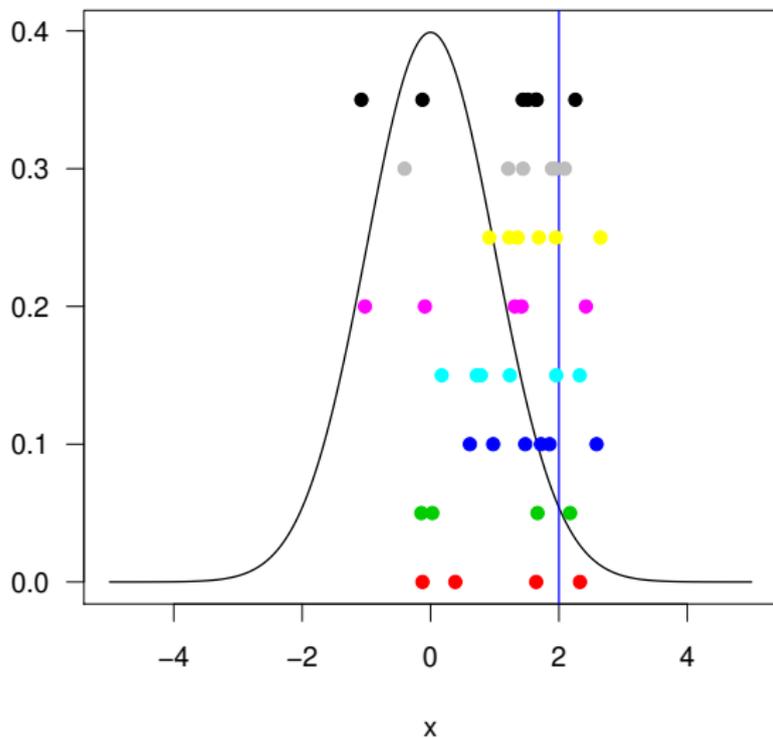
$$\hat{p} = (1 - 1/N) \{\text{nombre total de déplacements}\}$$



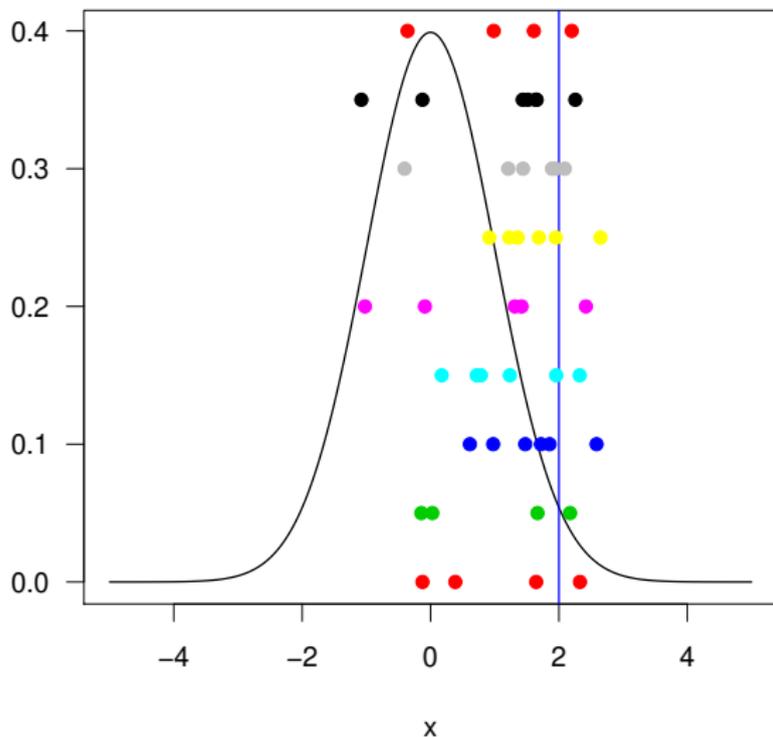
$$\hat{p} = (1 - 1/N) \{\text{nombre total de déplacements}\}$$



$$\hat{p} = (1 - 1/N) \{\text{nombre total de déplacements}\}$$



$$\hat{p} = (1 - 1/N)^{\{\text{nombre total de déplacements}\}} \rightarrow \hat{p} = (1 - 1/9)^{39} \approx 1,1 \cdot 10^{-2}$$



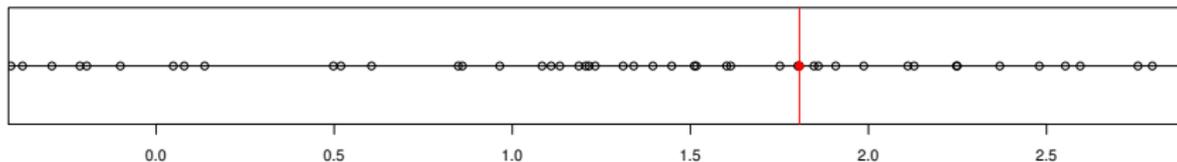
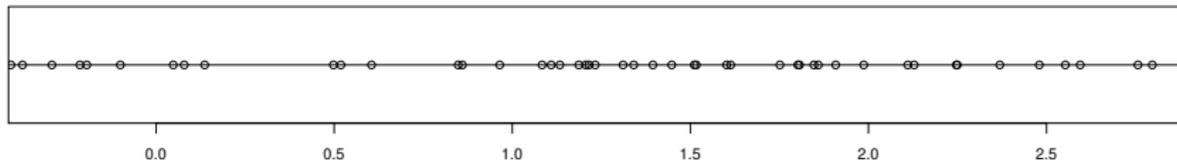
Concept [10]

Approcher un temps t quelconque par un/des temps des évènements d'un processus de Poisson

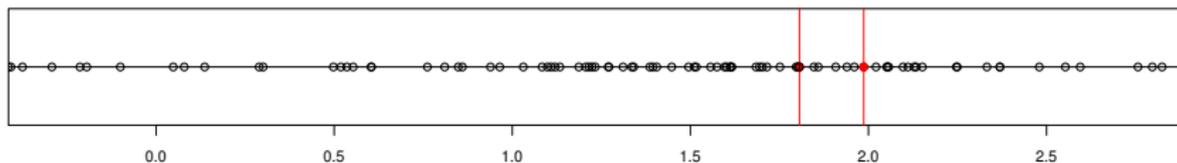
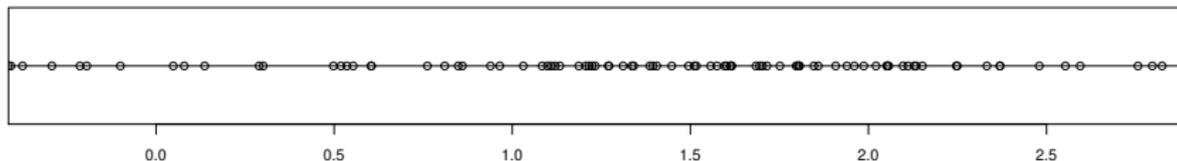
Simuler N processus pour *densifier la discrétisation* de $[0; +\infty[$

- Le temps t tombe *uniformément* dans l'intervalle $[T_{M_q}; T_{M_{q+1}}[$
- $\hat{q} = \frac{1}{2} (Y_m + Y_{m+1})$ avec $m = \lfloor E[M_q] \rfloor = \lfloor -N \log p \rfloor$
- Théorème Central Limite : $\sqrt{N} (\hat{q} - q) \xrightarrow[m \rightarrow \infty]{\mathcal{L}} \mathcal{N} \left(0, \frac{-p^2 \log p}{f(q)^2} \right)$
- Bornes sur le biais en $O(1/N)$

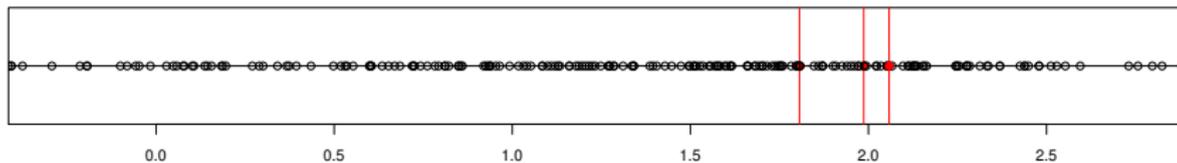
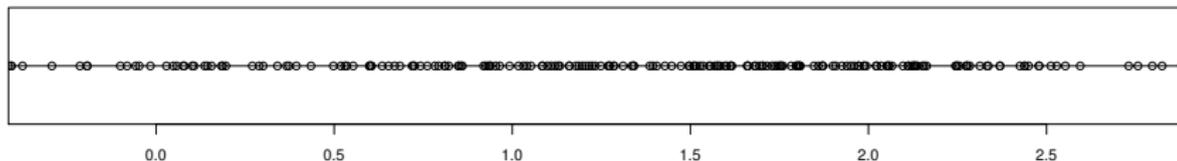
$$N = 10 \longrightarrow \lfloor -N \log p \rfloor = 37$$



$$N = 20 \longrightarrow \lfloor -N \log p \rfloor = 76$$



$$N = 50 \longrightarrow \lfloor -N \log p \rfloor = 190$$



- 1 Aperçu des différentes stratégies
 - Méthodes de Monte-Carlo
 - Méthodes avec méta-modèle
- 2 Simulation d'évènements rares
 - Processus ponctuel
 - Estimation de probabilité
 - Estimation de quantile
- 3 Cas pratique
- 4 Conclusion

Exemple de Der Kiureghian et De Stefano ([5, 6]) : oscillateur à deux degrés de liberté (dimension 8)

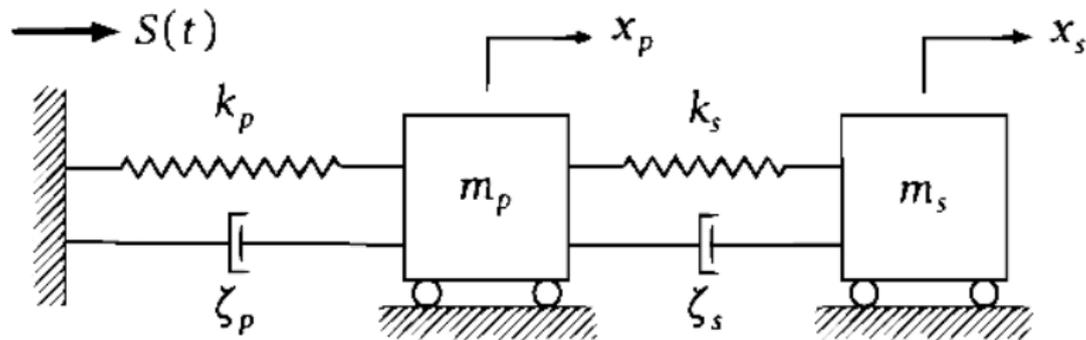


FIGURE: Un oscillateur à deux degrés de liberté soumis à une accélération de type bruit blanc

$$g(\mathbf{x}) = F_s - p k_s \sqrt{\mathbb{E}[x_s^2]}$$

avec F_s la capacité du second chariot et p un facteur de proportionnalité.

On regroupe les chaînes de Markov par *paquet* de $N = 20$ pour utiliser des algorithmes de type Metropolis-Hastings pour les générations conditionnelles.

On suppose que l'on dispose de $n_c = 50$ cœurs $\Rightarrow N \cdot n_c = 1000$ chaînes simulées

Monte-Carlo brut : valeur de référence ; résultats sur 100 simulations :

$E[F_s]$	Méthode	Proba	CV (%)	Nbr appels	Temps
15	MC	$4,80 \cdot 10^{-3}$	1	2.000.000	40.000
	PP	$4,80 \cdot 10^{-3}$	8,0	107.805	2.676
21,5	MC	$4,34 \cdot 10^{-5}$	4,8	10.000.000	200.000
	PP	$4,39 \cdot 10^{-5}$	10,5	201.700	4.734
27,5	SS [1]	$3,75 \cdot 10^{-7}$	2,9	###	###
	PP	$3,75 \cdot 10^{-7}$	12,5	296.932	6.800

- 1 Aperçu des différentes stratégies
 - Méthodes de Monte-Carlo
 - Méthodes avec méta-modèle
- 2 Simulation d'évènements rares
 - Processus ponctuel
 - Estimation de probabilité
 - Estimation de quantile
- 3 Cas pratique
- 4 Conclusion

Avantages

- Utilisation similaire à du Monte-Carlo classique pour une variable aléatoire réelle
- Cette approche *ajoute un log* au terme $1/p$
- Donne l'implémentation optimale des Multilevel Splitting Methods [1, 3, 8]

Inconvénients

- Difficulté de simuler selon les lois conditionnelles
- Utilisation de chaînes de Markov pour approcher les lois augmente le nombre d'appels



S-K Au and J L Beck.

Estimation of small failure probabilities in high dimensions by subset simulation.

Probabilistic Engineering Mechanics, 16(4) :263–277, 2001.



J-M Bourinet, F Deheeger, and M Lemaire.

Assessing small failure probabilities by combined subset simulation and support vector machines.

Structural Safety, 33(6) :343–353, 2011.



F Cérou, P Del Moral, T Furon, and A Guyader.

Sequential monte carlo for rare event estimation.

Statistics and Computing, 22(3) :795–808, 2012.



P Del Moral, A Doucet, and A Jasra.

Sequential monte carlo samplers.

Journal of the Royal Statistical Society : Series B (Statistical Methodology), 68(3) :411–436, 2006.



A Der Kiureghian and M De Stefano.

Efficient algorithm for second-order reliability analysis.

Journal of engineering mechanics, 117(12) :2904–2923, 1991.



V Dubourg, F Deheeger, and B Sudret.

Metamodel-based importance sampling for the simulation of rare events.

arXiv preprint arXiv :1104.3476, 2011.



B Echard, N Gayton, and M Lemaire.

Ak-mcs : An active learning reliability method combining kriging and monte carlo simulation.

Structural Safety, 33(2) :145–154, 2011.



A Guyader, N Hengartner, and E Matzner-Løber.

Simulation and estimation of extreme quantiles and extreme probabilities.

Applied Mathematics & Optimization, 64(2) :171–196, 2011.



T Igusa and A Der Kiureghian.

Dynamic characterization of two-degree-of-freedom equipment-structure systems.

Journal of engineering mechanics, 111(1) :1–19, 1985.



C Walter.

Moving particles : a parallel optimal multilevel splitting method with application in quantile estimation and metamodel based algorithm.

arXiv preprint arXiv :1405.2800, 2014.

Merci !

Commissariat à l'énergie atomique et aux énergies alternatives
CEA, DAM, DIF, F-91297 Arpajon, France

CEA
DAM
DIF