

## Estimation de probabilités et de quantiles rares pour la caractérisation d'une zone de retombée d'un engin

**Title:** Estimation of rare event probabilities and quantiles applied to an aerospace vehicle fall-back zone analysis

Jérôme Morio<sup>1</sup>, Rudy Pastel<sup>1,2</sup> et François Le Gland<sup>2</sup>

**Résumé :** Afin de quantifier les risques ainsi qu'évaluer des performances de système, il est souvent nécessaire d'estimer des quantiles et des probabilités faibles. Les techniques habituelles d'estimation de type méthode de Monte Carlo n'étant plus efficaces, nous détaillons les principales techniques d'estimation de probabilités rares telles que l'importance sampling, l'importance splitting ou la théorie des valeurs extrêmes. Ces différents algorithmes sont appliqués au cas de l'estimation d'une zone de retombée d'un engin spatial.

**Abstract:** Estimating rare event probability and quantile with a valuable accuracy is an important source of interest in reliability and safety. Since usual estimation techniques such as Monte Carlo method are not efficient for low probabilities, different methods have been investigated: importance splitting, importance sampling or extreme value theory. These algorithms are compared and then applied to the safety zone estimation of an aerospace vehicle.

**Mots-clés :** Estimation d'évènements rares, Théorie des valeurs extrêmes, Importance sampling, Importance splitting

**Keywords:** Rare event estimation, Extreme value theory, Importance sampling, Importance splitting

**Classification AMS 2000 :** 60-02, 65C05, 65C35, 62P30, 62-02, 62G32, 62G09

### 1. Introduction

L'évaluation de performances de système nécessite de plus en plus des techniques adaptées d'estimation. En effet, dans de nombreux secteurs tels que la défense, le spatial, mais également la finance, on tend à caractériser et par conséquent à probabiliser les risques inhérents à ces applications. Les domaines d'application concernant la sûreté de fonctionnement sont ainsi très étendus. Les études sur l'estimation d'évènements rares permettent, par exemple, de répondre à des questions telles que : quelle est la probabilité qu'un missile retombe à plus de  $x$  mètres de sa cible ? Comment déterminer la zone de retombée à 99% d'un étage de lanceur spatial ? Quelle est la probabilité de collision entre deux avions dans l'espace aérien ? Mais pour répondre à ces interrogations, les méthodes d'estimation et d'intégration les plus couramment utilisées (Monte Carlo, Quasi Monte Carlo [36]...) n'aboutissent pas à des résultats suffisamment précis. En effet, le nombre d'échantillons disponibles est souvent insuffisant pour calculer de manière

---

1. Onera - The French Aerospace Lab, F-91761 Palaiseau, France  
E-mail : [jerome.morio@onera.fr](mailto:jerome.morio@onera.fr) et E-mail : [rudy.pastel@onera.fr](mailto:rudy.pastel@onera.fr)

2. INRIA Rennes, ASPI Applications of interacting particle systems to statistics, Campus de Beaulieu 35042 Rennes France  
E-mail : [legland@irisa.fr](mailto:legland@irisa.fr)

précise des probabilités aussi faibles que celles que l'on souhaite estimer. Il est donc nécessaire de mettre en place des techniques adaptées à l'estimation de ces probabilités. Celles-ci peuvent être divisées principalement en deux grandes catégories : les techniques de paramétrisation des queues de densité de probabilité (PDF) basées sur la théorie des valeurs extrêmes, et les techniques de simulation et d'échantillonnage d'importance de types Importance Sampling ou Importance Splitting. Les techniques consacrées à l'exploitation de métamodèle ne seront pas développées dans cet article mais évoquées dans la conclusion.

Dans cet article, nous proposons donc de passer en revue les fondamentaux de ces techniques. Nous introduirons dans un premier temps la modélisation considérée tout au long de l'article, un système boîte noire déterministe composée d'entrées-sorties aléatoires, et nous comparons les performances de ces différentes techniques d'estimation dans ce contexte. Dans un second temps, nous appliquerons et confronterons ces mêmes techniques dans le cadre de l'estimation d'une zone de retombée d'un engin. La comparaison de ces différents algorithmes permettra ainsi d'en déduire leurs principaux domaines d'utilisation.

## 2. Contexte et problématique

### 2.1. Modélisation

Nous allons considérer dans cet article un modèle entrées-sorties de type boîte noire. Celui-ci, très général, est bien représentatif des modèles utilisés dans le cadre de l'ONERA (chaîne de calcul, système...). Les entrées  $X = (x_1, x_2, \dots, x_d)$  de ce système sont supposées  $d$ -dimensionnelles et aléatoires, et leur PDF jointe est notée  $f$ . Le système boîte noire possède une sortie monodimensionnelle  $Y$  définie par  $Y = \phi(X)$  avec  $\phi$  une fonction déterministe de  $\mathbb{R}^d \rightarrow \mathbb{R}$ . La sortie  $Y$  est donc par nature aléatoire et sa PDF, notée  $g$ , dépend notamment de la PDF jointe des entrées  $X$ .

Nous allons étudier dans cet article le cas général de l'estimation de la probabilité de dépassement d'un seuil  $S \in \mathbb{R}$  par la variable  $Y$ . Cela correspond donc à estimer l'intégrale suivante :

$$I_\phi = \int_S^{+\infty} g(y)dy = \int_{\mathbb{R}^d} \mathbf{1}_{\phi(x) > S} f(x) dx \quad (1)$$

Le terme  $\mathbf{1}_{\phi(x) > S}$  prend la valeur 1 si  $\phi(x) > S$  et 0 sinon.

### 2.2. Estimation de zone de retombée

La sauvegarde d'engins aérospatiaux (lanceurs, missiles,...) est une problématique importante dans le domaine de l'aéronautique car les conséquences d'une erreur d'estimation peuvent s'avérer dramatiques pour les populations [34]. Nous avons considéré un modèle multiparamétrique qui permet de déterminer les retombées aléatoires d'un engin sur le sol. Le poids de l'engin est d'environ 30 tonnes, l'altitude de largage est de 70 kilomètres et sa pente est de  $-10$  degrés. Les entrées suivantes peuvent influencer sur la position de retombée de l'engin :

- 1) l'erreur d'estimation en orientation.
- 2) l'erreur d'estimation en position et en vitesse due à la centrale inertielle.
- 3) les variations de la vitesse du vent.

entrées	densité
erreur d'orientation en degrés	$N(0, 5)$
erreur de position en $m$	$N(0, 1)$
erreur de vitesse en $m.s^{-1}$	$N(0, 5)$
erreur sur la masse en $kg$	$N(0, 15)$
erreur de pente en degrés	$N(0, 5)$
erreur sur la vitesse du vent en $m.s^{-1}$	$N(0, 3)$

TABLEAU 1. entrées du modèle de retombée  $(N(a, b))$  : densité gaussienne avec  $a$  pour moyenne et  $b$  pour écart-type).

4) la méconnaissance du poids de l'engin.

5) la méconnaissance de l'angle entre l'axe vertical et le vecteur vitesse.

L'ensemble de ces paramètres est résumé dans le tableau 1 avec les PDF associées. Le modèle étudié est donc une boîte noire  $\phi$  ayant 6 entrées que nous avons supposées gaussiennes  $X = (x_1, x_2, x_3, x_4, x_5, x_6)$  et 2 sorties  $Y$  et  $Z$  définissant la position de la retombée par rapport à une position de référence donnée par le point de retombée sans perturbation  $\phi(0, 0, 0, 0, 0, 0)$ . Pour simplifier le problème et éviter l'estimation de quantiles multidimensionnels [31], nous considérons par la suite uniquement la distance  $r$  par rapport à la position de référence  $r = \sqrt{Z^2 + Y^2}$ . Le modèle mathématique est donc le suivant :

$$r = \phi(x_1, x_2, x_3, x_4, x_5, x_6) \quad (2)$$

avec  $\phi$  une fonction continue  $\phi : \mathbb{R}^6 \rightarrow \mathbb{R}$ . Les figures 1 et 2 présentent les positions d'impact obtenues par simulations de Monte Carlo dans un repère centré sur  $\phi(0, 0, 0, 0, 0, 0)$ .

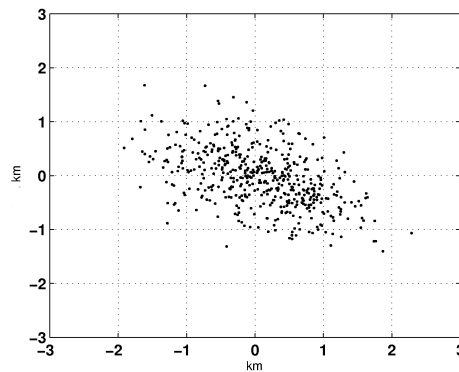


FIGURE 1. Positions d'impact de 500 simulations Monte Carlo.

### 2.3. Estimateur Monte Carlo

L'intégrale  $I_\phi$  peut être estimée de manière simple et efficace à l'aide de techniques de type Monte Carlo [49, 42, 29, 30, 43]. L'estimateur Monte Carlo de l'intégrale  $I_\phi$  est simplement

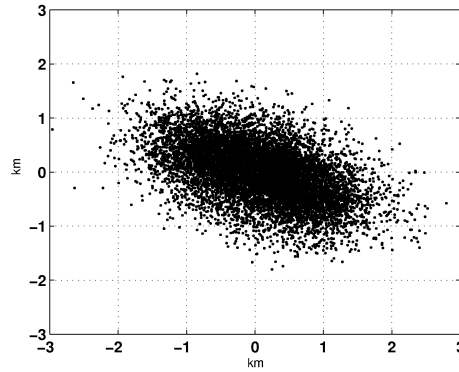


FIGURE 2. Positions d'impact de  $10^4$  simulations Monte Carlo.

donné par la moyenne empirique statistique :

$$I_{\phi}^{MC} = \frac{1}{N} \sum_{i=1}^N \mathbf{1}_{\phi(X_i) > s} \quad (3)$$

où  $X_1, \dots, X_N$  sont des échantillons aléatoires indépendants et identiquement générés selon la PDF  $f$ .

La précision de l'estimateur  $I_{\phi}^{MC}$  peut être analysée en terme de variance, laquelle dépend de  $N$  et de la valeur exacte  $I_{\phi}$ . Cette variance est donnée par :

$$\text{Var}_f(I_{\phi}^{MC}) = \frac{\text{Var}_f(\mathbf{1}_{\phi(X) > s})}{N} \quad (4)$$

d'où une précision en  $\frac{1}{\sqrt{N}}$ , ce qui signifie qu'il faut quadrupler le nombre  $N$  d'échantillons pour diviser par deux l'erreur d'estimation. Par définition de la variance, il est possible d'obtenir l'équation suivante :

$$\text{Var}_f(I_{\phi}^{MC}) = \frac{1}{N} \left( \int_{\mathbb{R}^d} \mathbf{1}_{\phi(x) > s}^2 f(x) dx - I_{\phi}^2 \right) \quad (5)$$

En remarquant que  $\mathbf{1}_{\phi(x) > s}^2 = \mathbf{1}_{\phi(x) > s}$  quelle que soit la valeur de  $x$ , on obtient alors :

$$\int_{\mathbb{R}^d} \mathbf{1}_{\phi(x) > s}^2 f(x) dx = \int_{\mathbb{R}^d} \mathbf{1}_{\phi(x) > s} f(x) dx = I_{\phi} \quad (6)$$

Puis, en remplaçant ce résultat dans l'équation (5), la variance de  $I_{\phi}^{MC}$  est donc définie par :

$$\text{Var}_f(I_{\phi}^{MC}) = \frac{1}{N} (I_{\phi} - I_{\phi}^2) \quad (7)$$

L'écart-type de  $I_{\phi}^{MC}$ ,  $\sigma_{I_{\phi}^{MC}}$ , est défini à partir de la variance

$$\sigma_{I_{\phi}^{MC}} = \sqrt{\text{Var}_f(I_{\phi}^{MC})} = \frac{1}{\sqrt{N}} \sqrt{I_{\phi} - I_{\phi}^2} \quad (8)$$

La convergence en  $\frac{1}{\sqrt{N}}$  de l'estimateur  $I_\phi^{MC}$  est donc bien vérifiée. Néanmoins, lorsqu'il s'agit d'estimer des évènements de faible probabilité, c'est-à-dire quand  $I_\phi$  prend de faibles valeurs, l'écart-type de  $I_\phi$  à taille d'échantillons  $N$  fixée est de l'ordre  $\sqrt{\frac{I_\phi}{N}}$ . Dans ce cas, l'écart-type  $\sqrt{\frac{I_\phi}{N}}$  est bien plus élevé que l'estimateur  $I_\phi$ , sauf si le nombre d'échantillons  $N$  prend de fortes valeurs. Cela est souvent problématique en réalité. En effet, le coût et le temps de calcul des échantillons ne permettent généralement pas d'utiliser des valeurs de  $N$  trop importantes. Cependant, la parallélisation sur de multiples processeurs augmente la capacité de calcul et permet ainsi de résoudre ce problème. Utiliser l'estimateur  $I_\phi^{MC}$  est donc particulièrement imprécis car l'écart-type est au moins du même ordre de grandeur que l'estimation. De plus, dans le cadre des évènements rares, l'écart-type n'est pas forcément adapté à la mesure de la précision d'une estimation. Il est souvent plus intéressant de définir un écart-type relatif. Dans le cas présent, l'écart-type relatif de l'estimateur  $I_\phi^{MC}$  est donné par le rapport  $\frac{\sigma_{I_\phi^{MC}}}{I_\phi^{MC}}$ . Quand  $I_\phi$  tend vers zéro, on obtient alors :

$$\lim_{I_\phi \rightarrow 0} \frac{\sigma_{I_\phi^{MC}}}{I_\phi^{MC}} = \lim_{I_\phi \rightarrow 0} \frac{1}{\sqrt{NI_\phi}} = +\infty \quad (9)$$

Si  $I_\phi = 10^{-6}$  et que l'on souhaite  $\frac{\sigma_{I_\phi^{MC}}}{I_\phi^{MC}} = 10^{-1}$ , il faut alors  $N = 10^8$  échantillons.  $I_\phi^{MC}$  ne peut donc pas être utilisé avec une bonne précision pour l'estimation d'évènements rares. D'autres techniques d'estimation plus adaptées à des probabilités  $I_\phi$  faibles doivent être mises en place. A titre d'exemple, le tableau 2 présente l'estimation de la probabilité  $P(Y > S) = I_S$ , en supposant  $Y$  une variable aléatoire gaussienne centrée réduite. La valeur théorique des probabilités est aussi présentée dans ce tableau. L'augmentation du nombre d'échantillons ne permet qu'une réduction lente de l'écart-type relatif d'estimation. Pour estimer précisément toutes ces probabilités, il faudrait plus de  $10^9$  échantillons, ce qui peut s'avérer problématique dans un cadre d'application réaliste. De plus, l'aspect cyclique de certains générateurs de nombres aléatoires ne permet plus une estimation fiable.

L'estimation de quantiles de distribution de probabilité est aussi un facteur important dans la maîtrise des risques. Le  $\alpha$ -quantile de la distribution de la variable aléatoire  $Y$  peut être défini comme la valeur  $q_\alpha$  telle que :

$$q_\alpha = \inf \{y : G(y) \geq \alpha\}$$

La valeur de  $q_\alpha$  peut être estimée par Monte Carlo [48, 42], mais le résultat n'est pas suffisamment précis pour des quantiles rares. La méthode de Monte Carlo classique pour l'estimation de quantile s'effectue de la manière suivante : simuler un échantillon  $X_1, \dots, X_N$  selon la densité  $f$ , appliquer la fonction  $\phi$  et poser  $Y_1 = \phi(X_1), \dots, Y_N = \phi(X_N)$ , ré-ordonner dans le sens croissant l'échantillon obtenu, de sorte que  $Y_{(1)} \leq \dots \leq Y_{(N)}$ . L'estimateur Monte Carlo du quantile  $q_\alpha$  est alors défini par

$$q_\alpha^{MC} = \phi(X_{[\alpha N]}) = Y_{[\alpha N]} \quad (10)$$

où  $[z]$  désigne la partie entière de  $z$ . Le biais est d'ordre  $O(1/N)$  et si la fonction de répartition de  $G_Y$  de la variable aléatoire  $Y = \phi(X)$  admet une dérivée  $g$  non nulle au point  $q_\alpha$ , alors on peut montrer que la variance de l'estimateur est de la forme :

$$\text{Var}_f(q_\alpha^{MC}) \approx \frac{\alpha(1-\alpha)}{Ng^2(q_\alpha)} \quad (11)$$

N	$I_2^{MC}$	$I_3^{MC}$	$I_4^{MC}$	$I_5^{MC}$
$10^3$	0.0230 (19%)	0.0120 (78%)	$2 \cdot 10^{-5}$ (700%)	0 (?%)
$10^4$	0.0227 (6%)	0.0135 (26%)	$3.5 \cdot 10^{-5}$ (154%)	0 (?%) %
$10^5$	0.0228 (2%)	0.0133 (8%)	$3.05 \cdot 10^{-5}$ (56%)	$9 \cdot 10^{-7}$ (319%)
$10^6$	0.0228 (0.5%)	0.0135 (2%)	$3.24 \cdot 10^{-5}$ (20%)	$3.1 \cdot 10^{-7}$ (187%)
$10^7$	0.0228 (0.2%)	0.0135 (0.8%)	$3.14 \cdot 10^{-5}$ (4%)	$3.28 \cdot 10^{-7}$ (47%)
théorie	0.0228	0.0135	$3.17 \cdot 10^{-5}$	$2.86 \cdot 10^{-7}$

TABEAU 2. Estimation moyenne de  $I_S^{MC}$  et des erreurs relatives d'estimation par méthode Monte Carlo pour différentes valeurs de  $N$  et de  $S$  répétée 100 fois.

N	$q_{0.95}^{MC}$	$q_{0.999}^{MC}$	$q_{0.99999}^{MC}$	$q_{0.9999999}^{MC}$
$10^3$	1.63 (3%)	3.11 (9.3%)	3.28 (11%)	3.28 (11%)
$10^4$	1.65 (1.2%)	3.09 (2.8%)	3.86 (8.8%)	3.86 (8.8%)
$10^5$	1.64 (0.3%)	3.09 (1%)	4.26 (4.5%)	4.36 (5.6%)
$10^6$	1.65 (0.1%)	3.09 (0.3%)	4.26 (1.6%)	5.08 (4.9%)
$10^7$	1.64 (0.04%)	3.09 (0.09%)	4.26 (0.5%)	5.22 (3.2%)
théorie	1.65	3.09	4.27	5.20

TABEAU 3. Estimation moyenne de quantiles rares d'une variable gaussienne centrée réduite par méthode Monte Carlo pour différentes valeurs de  $N$  répétée 100 fois.

pour  $N$  grand. En supposant  $Y$  une variable aléatoire gaussienne centrée réduite, le tableau 3 présente l'estimation de différents quantiles rares de la variable  $Y$  par méthode Monte Carlo pour différentes tailles d'échantillons. A l'instar des probabilités, l'estimation de quantiles rares ne peut être effectuée à l'aide des méthodes usuelles comme Monte Carlo.

Dans le cadre de la retombée d'un engin, nous avons estimé les probabilités de retombée à différentes distances de la position prévue par la méthode Monte Carlo, ainsi que l'écart-type relatif d'estimation. Ces résultats sont récapitulés dans le tableau 4. Lorsque les probabilités estimées deviennent trop faibles, la méthode Monte Carlo n'est plus adaptée. Les mêmes conclusions que dans le cas d'une variable gaussienne centrée réduite peuvent être tirées. Il est donc nécessaire de mettre en place des méthodes adaptées à l'estimation de probabilités et de quantiles rares.

Il existe deux grandes familles de méthodes pour caractériser des événements rares :

- les techniques de paramétrisation de la queue de distribution qui reposent sur la théorie des valeurs extrêmes.
- les techniques de simulation basées sur une génération astucieuse des échantillons. Parmi

seuil $S$ en km	$I_\phi^{MC}$	écart-type relatif
0.5	0.72	0.64%
1	0.29	1.4%
1.5	0.074	3.2%
2	0.0125	8.1%
2.5	0.0013	26.1%
3	$4.6 \cdot 10^{-5}$	132.6%
3.5	$1.0 \cdot 10^{-6}$	1000%
4	?	?

TABEAU 4. Estimation moyenne de  $I_\phi^{MC}$  avec  $N = 10000$  simulations Monte Carlo répétée 100 fois.

ces techniques, il en existe principalement deux, que sont l'Importance Sampling (= échantillonnage pondéré ou préférentiel) et l'Importance Splitting.

Dans la suite de cet article, nous allons détailler ces différentes méthodes et apprécier dans quel cadre elles peuvent être appliquées.

### 3. Théorie des valeurs extrêmes

La théorie des valeurs extrêmes est particulièrement utilisée en finance et en climatologie pour estimer des risques (produits financiers, inondations...). L'objectif de cette méthode est d'obtenir, à l'aide d'échantillons d'une variable aléatoire, une paramétrisation de sa queue de distribution. Elle permet donc de déduire une estimation des quantiles et des probabilités rares [15, 40]. Par ailleurs, la théorie des valeurs extrêmes se rapproche du théorème central limite qui caractérise l'évolution de la moyenne d'une somme de variables aléatoires indépendantes. La théorie des valeurs extrêmes permet quant à elle, de caractériser l'évolution de ces maxima.

#### 3.1. Principe

Soient  $(Y_1, Y_2, \dots, Y_N)$  une série d'échantillons aléatoires indépendants et identiquement distribués (i.i.d.) suivant la PDF de la variable de sortie du système boîte noire  $Y$ . Soit  $G_Y$  la fonction de répartition de  $Y$  et  $M_N = \max(Y_1, Y_2, \dots, Y_N)$  le maximum des observations. L'évolution de la loi des maxima peut s'obtenir de la manière suivante :

$$\begin{aligned} P(M_N \leq y) &= P(\max(Y_1, Y_2, \dots, Y_N) \leq y) \\ &= P(Y_1 \leq y, Y_2 \leq y, \dots, Y_N \leq y) = (G_Y(y))^N \end{aligned}$$

Dans le cas d'un échantillon i.i.d, la loi des maxima s'obtient donc facilement dès lors que  $G_Y$  est connue. Le résultat fondamental de la théorie des valeurs extrêmes est que, sous des conditions de régularité très générales, la loi des extrêmes appartient à une famille composée de trois lois de probabilités [14, 23, 24, 11, 13]. S'il existe deux suites de nombres réels  $(a_N, b_N)$ , avec  $a_N > 0$  et une fonction de distribution  $H$  non dégénérée telles que :

$$P\left(\frac{M_N - b_N}{a_N} \leq t\right) = G^N(a_N t + b_N) \rightarrow H(t), \quad N \rightarrow +\infty \quad (12)$$

pour tout  $t \in \mathbb{R}$ , alors  $H$  ne peut être qu'une distribution de type Fréchet, Gumbel ou Weibull négative. Ces trois lois caractérisent chacune un comportement spécifique des distributions extrêmes. On les distingue notamment au travers de la vitesse de décroissance de la probabilité des grandes valeurs et de l'étendue de variation des données. Si la décroissance est rapide, par exemple de type exponentielle, alors la loi des extrêmes est la loi de Gumbel. Si au contraire, elle est plus lente, de type puissance, on se réfère alors à l'étendue de variation des données. Si celle-ci est bornée, alors la loi des extrêmes est la loi de Weibull, et si ce n'est pas le cas, c'est la loi de Fréchet.

Ces trois distributions peuvent en fait se regrouper en une seule famille paramétrique : la distribution généralisée des valeurs extrêmes (Generalized Extreme Value = GEV en anglais). La distribution GEV se définit de la manière suivante :

$$H(t) = \exp\left(-\left[1 + \varepsilon \frac{t - \mu}{\sigma}\right]^{-\frac{1}{\varepsilon}}\right)$$

où  $(\mu, \sigma, \varepsilon)$  sont respectivement les paramètres de localisation (lié à la valeur la plus probable de la loi), d'échelle (paramètre de dispersion) et de forme (plus cet indice est élevé en valeur absolue, plus le poids des extrêmes dans la distribution initiale est important) de la distribution GEV avec  $\sigma > 0$ . Les distributions de Fréchet, Gumbel ou Weibull négative sont alors obtenues pour des cas particuliers de  $(\mu, \sigma, \varepsilon)$ . On caractérise ainsi la densité de probabilité du maximum d'un échantillon d'une variable aléatoire. Néanmoins, pour un jeu de données réelles, on imagine bien que la détermination de la GEV correspondante n'est pas évidente. En effet, il faut pour cela séparer en blocs aléatoires (appelés blocs maxima), les données d'entrées et déterminer les valeurs maxima de chacun des blocs. La densité GEV est alors estimée à partir de l'échantillon de ces valeurs maxima. Différentes méthodes pour estimer les paramètres de la densité GEV à partir d'un jeu d'échantillons existent mais ne seront pas traitées dans cet article.

### 3.2. Dépassement de seuil

Le problème qui nous intéresse dans cet article correspond à l'estimation d'une probabilité de dépassement de seuil [10, 28, 39, 8]. A partir des résultats précédents sur la distribution GEV, il est tout à fait possible de résoudre ce problème. L'approche POT (Peaks Over Threshold) est la plus utilisée dans les applications que l'on peut trouver dans la littérature des valeurs extrêmes. Soit  $\mu_{end} = \sup(y \in \mathbb{R} | G_Y(y) < 1) \leq \infty$ , la borne supérieure du support de  $G_Y$ . La densité conditionnelle de la variable  $Y$ , conditionnée à dépasser un seuil  $\mu$ , approche une densité de probabilité de Pareto généralisée (ou GPD, pour *Generalized Pareto Distribution*) quand  $\mu$  tend vers  $\mu_{end}$ . En effet, si  $Y$  vérifie l'équation (12), alors on peut en déduire :

$$P(Y \leq t | Y > \mu) \rightarrow H(t), \quad \mu \rightarrow \mu_{end}$$

$$H(t) = 1 - \left(1 + \varepsilon \frac{t - \mu}{\sigma}\right)^{-\frac{1}{\varepsilon}}$$

où  $(\mu, \sigma, \varepsilon)$  sont respectivement les paramètres de localisation, d'échelle et de forme de la distribution GPD avec  $\sigma > 0$ . Si l'on connaît  $H(t)$  et  $P(Y > \mu)$  pour la série  $(Y_1, Y_2, \dots, Y_N)$ , on est alors capable d'estimer les quantiles et les probabilités de la variable aléatoire  $Y$  au delà du seuil  $\mu$ . A partir des données, il faut donc estimer les différents paramètres de la GPD  $(\mu, \sigma, \varepsilon)$ . Pour une série d'échantillons donnés  $(Y_1, Y_2, \dots, Y_N)$ , il faut tout d'abord choisir la valeur du seuil  $\mu$ . C'est une étape capitale du processus d'estimation, car elle permet de déterminer quels échantillons seront utilisés pour l'estimation de  $\sigma$  et  $\varepsilon$ . Ce choix est d'autant plus important qu'il faut effectuer un compromis entre biais et variance. En effet, si  $\mu$  est trop faible, c'est-à-dire si l'on prend beaucoup d'échantillons  $(Y_1, Y_2, \dots, Y_N)$  pour l'estimation, un biais apparaît dû à l'influence du comportement central de la PDF. A l'inverse, si  $\mu$  est trop élevé, la variance d'estimation est forte car peu de données sont prises en compte dans l'estimation. Il existe de nombreuses



méthodes pour estimer le seuil  $\mu$  de manière optimale. Les méthodes les plus connues sont basées sur l'estimateur de Hill et sur la moyenne empirique des dépassements. De manière générale, ces méthodes sont empiriques et aucune d'entre elles n'est reconnue unanimement comme étant la meilleure. Ce choix étant important dans la caractérisation de la GPD, il peut souvent être judicieux d'utiliser plusieurs estimateurs et de comparer leurs valeurs.

Connaissant la valeur estimée  $\hat{\mu}$  de la GPD, il est alors possible d'estimer les paramètres  $\sigma$  et  $\varepsilon$ . De nombreuses méthodes existent et l'estimateur du maximum de vraisemblance est très intéressant en pratique. On le privilégiera par rapport à d'autres estimateurs basés sur la méthode des moments [21].

### 3.3. Application sur un cas simple

Considérons le cas de l'estimation de  $I_S$  par la théorie des valeurs extrêmes avec  $Y$  une densité de probabilité gaussienne centrée réduite et  $S$  un seuil. Le paramètre  $\mu$  de la GPD est déterminé par la méthode graphique issue de la moyenne empirique des dépassements et les paramètres  $\sigma$  et  $\varepsilon$  par maximum de vraisemblance. La figure 3 présente la densité GPD estimée à partir de 10000 échantillons gaussiens centrés réduits. Les courbes GPD et gaussienne sont proches au niveau de la queue de la distribution, ce qui valide la démarche issue de la théorie des valeurs extrêmes. Le

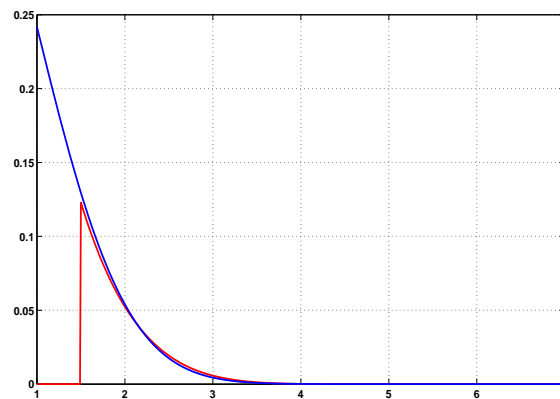


FIGURE 3. La densité GPD (rouge) estimée à l'aide de 10000 échantillons gaussiens centrés réduits et la densité gaussienne centrée réduite (bleu).

tableau 5 présente les résultats d'estimation obtenus à l'aide de  $N$  échantillons. La théorie des valeurs extrêmes permet d'approcher des probabilités de l'ordre de  $\frac{10^{-2}}{N}$  ce qui est une amélioration importante par rapport à une méthode de Monte Carlo. Le fait que l'estimation du paramètre  $\mu$  soit parfois empirique peut s'avérer problématique et cela peut avoir des conséquences non négligeables sur le résultat obtenu qui pourra être biaisé. Le tableau 6 présente aussi l'estimation de quantiles rares par théorie des valeurs extrêmes. La théorie des valeurs extrêmes permet donc d'évaluer des quantiles rares et sans rééchantillonnage ce qui est un atout non négligeable lorsque l'on ne possède pas le code de calcul des données mais uniquement certains échantillons.

N	$I_2^{TVE}$	$I_3^{TVE}$	$I_4^{TVE}$	$I_5^{TVE}$
$10^3$	0.0214	0.00139	$1.87 \cdot 10^{-5}$	$3.75 \cdot 10^{-10}$
$10^4$	0.0227	0.00129	$1.08 \cdot 10^{-5}$	$8.94 \cdot 10^{-14}$
$5 \cdot 10^4$	0.0224	0.00137	$1.47 \cdot 10^{-5}$	$1.95 \cdot 10^{-11}$
$10^5$	?	0.00134	$3.76 \cdot 10^{-5}$	$2.92 \cdot 10^{-7}$
théorie	0.0228	0.00135	$3.17 \cdot 10^{-5}$	$2.86 \cdot 10^{-7}$

TABEAU 5. Estimation de probabilités de dépassement par la théorie des valeurs extrêmes pour différentes valeurs de  $N$ .

N	$q_{0.95}^{TVE}$	$q_{0.99}^{TVE}$	$q_{0.999}^{TVE}$	$q_{0.99999}^{TVE}$	$q_{0.9999999}^{TVE}$
$10^3$	1.78 [1.68, 1.90]	2.31 [1.84, 2.58]	2.85 [2.57, 3.75]	3.44 [2.83, 4.09]	3.71 [2.87, 4.09]
$10^4$	?	2.58 [2.64, 2.71]	3.37 [2.65, 3.83]	4.42 [3.89, 5.62]	5.10 [4.14, 5.62]
$5 \cdot 10^4$	?	2.57 [2.56, 2.58]	3.26 [3.20, 3.45]	4.22 [3.97, 5.18]	4.79 [4.31, 6.30]
$10^5$	?	?	3.27 [2.99, 3.43]	4.31 [4.09, 5.31]	5.00 [4.51, 6.63]

TABEAU 6. Estimation de quantiles rares par la théorie des valeurs extrêmes pour différentes valeurs de  $N$  ainsi que leur intervalle de confiance à 95%.

Lorsque le nombre d'échantillons augmente, la valeur du seuil  $\mu$  augmente également. Certaines estimations des tableaux 5 et 6 pour de faibles seuils ou de faibles quantiles ne sont donc plus possibles (la valeur d'estimation est remplacée par un ? dans ces tableaux). Les erreurs relatives des estimations données par la théorie des valeurs extrêmes n'ont pas été déterminées.

### 3.4. Application sur l'estimation d'une zone de retombée

Dans le cadre de l'estimation de zones de retombée, un jeu de 10000 échantillons aléatoires de rayon de retombée a été réalisé par méthode de Monte Carlo. A partir de ces données, le paramètre  $\mu$  de la GPD a été déterminé par la méthode graphique issue de la moyenne empirique des dépassements et les paramètres  $\sigma$  et  $\varepsilon$  ont été déterminés par maximum de vraisemblance. La figure 4 présente l'estimation de la densité GPD correspondant à ces échantillons ainsi que la densité estimée par noyaux sur les échantillons. Il y a une bonne correspondance entre ces deux densités. La théorie des valeurs extrêmes permet alors d'aboutir aux résultats donnés par le tableau 7 qui donne un premier ordre de grandeur sur les probabilités en jeu pour les différents seuils de la retombée.

seuil $S$ en km	$I_\phi^{TVE}$
0.5	?
1	?
1.5	0.070
2	0.012
2.5	0.0016
3	$7.8 \cdot 10^{-5}$
3.5	$4.2 \cdot 10^{-6}$
4	$1.3 \cdot 10^{-7}$

TABEAU 7. Estimation de  $I_\phi^{TVE}$  avec  $N = 10000$  simulations utilisées pour estimer les paramètres GPD.

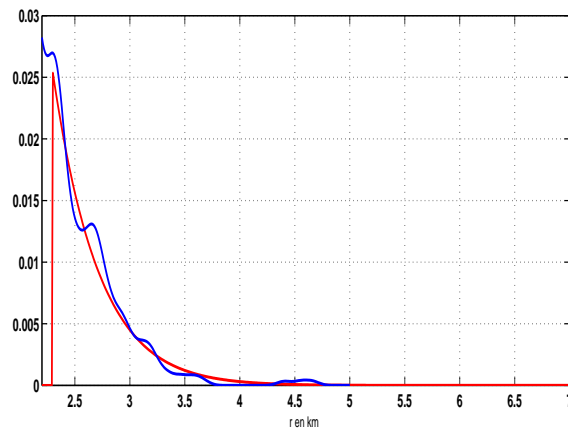


FIGURE 4. La densité GPD (rouge) estimée à l'aide de 10000 échantillons et la densité estimée par noyaux gaussiens (bleu) dans le cas de la zone de retombée.

### 3.5. Conclusion

La théorie des valeurs extrêmes permet de décrire de manière paramétrique la queue de distribution d'une variable aléatoire. D'un point de vue pratique, cette méthode présente tout d'abord l'avantage de pouvoir estimer sur un unique jeu de données, différents quantiles et différentes probabilités. Nous verrons par la suite que ce n'est pas le cas de toutes les méthodes d'estimation d'évènements rares. Néanmoins, il faut rester prudent sur les résultats obtenus par cette méthode. Tout d'abord, la théorie des valeurs extrêmes est basée sur des propriétés asymptotiques des PDF. En réalité, un nombre infini de données n'est jamais disponible. De plus, le caractère empirique du choix du paramètre de localisation influence de manière non négligeable les estimations. La théorie des valeurs extrêmes semble donc adaptée lorsque le nombre de données est faible, pour obtenir un ordre de grandeur des estimations. Ainsi, lorsqu'un rééchantillonnage ne peut être obtenu, il est nécessaire d'utiliser la théorie des valeurs extrêmes. Cependant, lorsque des valeurs de  $N$  plus élevées sont autorisées et disponibles, les méthodes d'échantillonnage que nous allons présenter par la suite semblent plus pertinentes.

## 4. Importance Sampling

La méthode d'Importance Sampling (IS) est la méthode la plus connue pour estimer des probabilités d'évènements rares [1, 2, 12]. Elle a, en effet, l'avantage d'être simple à mettre en place et permet de diminuer la variance de l'estimateur Monte Carlo à l'aide d'une densité de probabilité auxiliaire. La difficulté de l'algorithme consiste à optimiser cette densité auxiliaire pour diminuer efficacement la variance d'estimation.

#### 4.1. Principe

L'objectif de l'IS est de réduire la variance de l'estimateur Monte Carlo  $I_\phi^{MC}$  définie par l'équation (3) sans augmenter le nombre d'échantillons  $N$ . L'idée principale est de générer les échantillons  $X_1, \dots, X_N$  à l'aide d'une densité auxiliaire  $h$  plus génératrice en événements rares que la densité initiale  $f$ , puis d'introduire une pondération dans l'estimateur. Le terme  $I_\phi$  est ensuite estimé de la manière suivante :

$$I_\phi^{IS} = \frac{1}{N} \sum_{i=1}^N \mathbf{1}_{\phi(X_i) > S} \frac{f(X_i)}{h(X_i)} \quad (13)$$

L'estimateur  $I_\phi^{IS}$  est un estimateur non-biaisé de  $I_\phi$  car :

$$E_h(I_\phi^{IS}) = \int_{\mathbb{R}^d} \mathbf{1}_{\phi(x) > S} f(x) dx = I_\phi \quad (14)$$

avec  $E_h$  l'espérance mathématique par rapport à la loi  $h$ . La variance de  $I_\phi^{IS}$  est donnée par :

$$\text{Var}_h(I_\phi^{IS}) = \frac{\text{Var}_h(\mathbf{1}_{\phi(X) > S} w(X))}{N} \quad (15)$$

avec  $w(x) = \frac{f(x)}{h(x)}$ . On peut ainsi obtenir l'équation suivante :

$$\text{Var}_h(I_\phi^{IS}) = \frac{1}{N} \left( \int_{\mathbb{R}^d} \mathbf{1}_{\phi(x) > S} w(x)^2 h(x) dx - I_\phi^2 \right) = \frac{1}{N} (E_h(\mathbf{1}_{\phi(X) > S} w(X)^2) - I_\phi^2) \quad (16)$$

La variance de l'estimateur IS dépend donc notamment du choix de la PDF  $h$ . Si la PDF  $h$  est bien choisie, alors la variance de l'estimateur IS peut devenir très basse. A l'inverse, si  $h$  n'est pas adaptée à l'estimation, les résultats de l'Importance Sampling pourront même avoir une variance bien supérieure à l'estimateur Monte Carlo. La section suivante permet de définir la densité auxiliaire optimale de l'IS qui minimise la variance d'estimation.

#### 4.2. Densité auxiliaire optimale de l'IS

L'équation (16) décrit la variance de la technique IS. L'objectif de l'IS est de minimiser la variance d'estimation. C'est pourquoi on peut définir la densité auxiliaire optimale de l'IS comme la densité qui minimise la variance  $\text{Var}_h(I_\phi^{IS})$ . Les variances étant des quantités non négatives, la densité auxiliaire optimale de l'IS  $h_{opt}$  est donc déterminée en annulant la variance dans l'équation (15).

$$\frac{1}{N} \left( E_{h_{opt}} \left( \mathbf{1}_{\phi(X) > S} \frac{f(X)^2}{h_{opt}(X)^2} \right) - I_\phi^2 \right) = 0 \quad (17)$$

Une variable aléatoire de variance nulle est presque sûrement constante, donc égale à sa moyenne, ce qui permet d'obtenir facilement

$$\mathbf{1}_{\phi(X) > S} \frac{f(X)}{h_{opt}(X)} = I_\phi \quad (18)$$

Nous pouvons ainsi déterminer une expression de  $h_{opt}$  de la manière suivante :

$$h_{opt}(x) = \frac{\mathbf{1}_{\phi(x) > Sf(x)}}{I_\phi} \quad (19)$$

La PDF auxiliaire  $h_{opt}$  dépend malheureusement de  $I_\phi$  qui est la probabilité inconnue que l'on cherche à estimer. La PDF  $h_{opt}$  est par conséquent inutilisable en pratique. Il est nécessaire d'effectuer une optimisation afin de déterminer une PDF  $h$  de telle sorte qu'elle se rapproche de la densité  $h_{opt}$ . Il existe plusieurs méthodes pour approcher la densité auxiliaire optimale. Citons, sans exhaustivité, les méthodes suivantes qui optimisent la densité de l'IS :

- de manière paramétrique (Cross-Entropy, changement d'échelle, translation)
- de manière non paramétrique (Nonparametric Adaptive Importance Sampling)
- de manière semi-paramétrique (twisted densities)

Les sections suivantes consisteront donc à brièvement décrire ces méthodes d'optimisation de la densité auxiliaire de l'IS.

### 4.3. Optimisation de la densité auxiliaire

#### 4.3.1. Changement d'échelle et translation

La méthode par changement d'échelle consiste à supposer a priori que la PDF  $h$  est définie par le modèle suivant :

$$h(x) = \frac{1}{a^d} f\left(\frac{x}{a}\right) \quad (20)$$

avec  $a \in \mathbb{R}$ . Un bon choix de paramétrage permet de tirer dans la queue de la distribution. Il faut donc minimiser la variance de l'estimateur IS vis-à-vis du paramètre  $a$ . Cette méthode n'est néanmoins efficace que dans le cas où la dimension  $d$  n'est pas trop importante. La figure 5 présente la moyenne et l'écart-type d'estimation dans le cas simple de l'estimation de  $I_5$  avec  $Y$  une variable aléatoire gaussienne centrée réduite pour  $N = 1000$ . L'écart-type d'estimation atteint un minimum en  $a = 7.1$  et tend vers l'infini pour de fortes valeurs de  $a$ . Lorsque  $a$  est bien choisi, l'IS améliore fortement la précision de l'estimation.

Une autre méthode simple utilise la technique de translation et suppose que la densité auxiliaire de l'IS  $h$  est définie de la manière suivante :

$$h(x) = f(x - c) \quad (21)$$

avec  $c \in \mathbb{R}^d$ , le paramètre de la translation. Celui-ci doit être choisi de telle sorte que  $c$  minimise la variance. La méthode de translation est moins sensible à la dimensionnalité que la méthode par changement d'échelle et est utilisée dans différentes applications. La figure 6 présente la moyenne et l'écart-type d'estimation dans le cas simple de l'estimation de  $I_5$  avec  $Y$  une variable aléatoire gaussienne centrée réduite pour  $N = 1000$ . L'écart-type d'estimation atteint un minimum en  $c = 4.96$  ce qui était prévisible. Cette méthode est bien adaptée à ce type d'estimation car le positionnement du seuil est connu a priori. Ces méthodes sont globalement simples à mettre en place et permettent une évaluation pertinente de la probabilité dans les cas où la dimensionnalité est faible. Pour optimiser la densité auxiliaire de l'IS dans des cas plus difficiles, la méthode d'optimisation par Cross-Entropy peut être notamment utilisée.

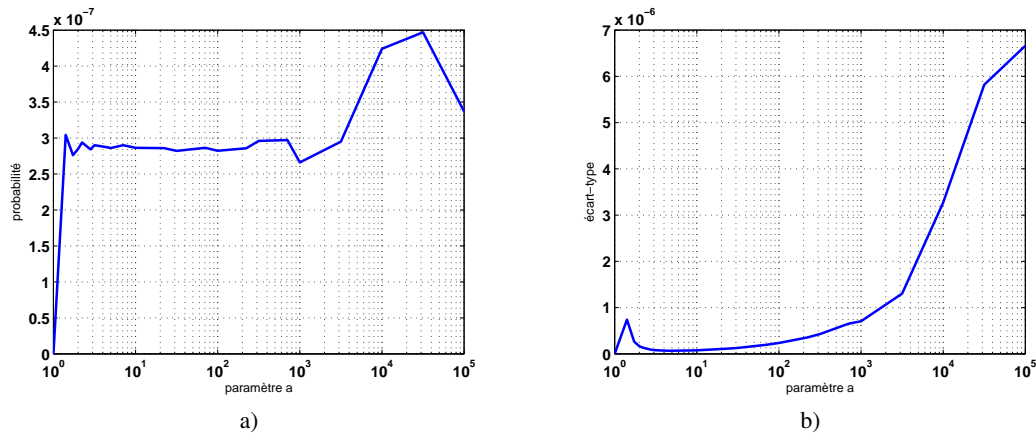


FIGURE 5. Influence du paramètre  $a$  de facteur d'échelle sur la moyenne a) et l'écart-type d'estimation b).

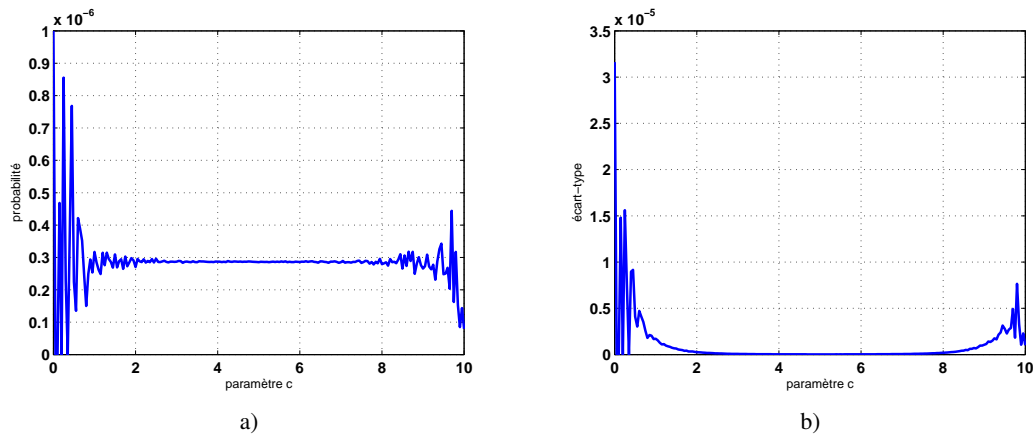


FIGURE 6. Influence du paramètre  $c$  de translation sur la moyenne a) et l'écart-type d'estimation b).

#### 4.3.2. Optimisation par Cross-Entropy

L'optimisation basée sur la Cross-Entropy (CE) ou entropie croisée permet d'approcher la densité auxiliaire optimale de l'IS parmi une famille de lois paramétriques (par exemple pour des densités gaussiennes, on cherche à optimiser la moyenne et la variance de la loi) [3, 20, 44, 45, 46] par rapport à un critère appelé entropie croisée. Soit  $h_\lambda$  une famille de densités de probabilité, paramétrée par le vecteur  $\lambda \in \Delta$  où  $\Delta$  est le domaine multidimensionnel de variation des paramètres de la densité. L'entropie croisée, appelée aussi distance de Kullback-Leibler  $D(q, p)$  entre deux densités de probabilité  $p$  et  $q$  est définie par :

$$D(q, p) = E_q \ln \left( \frac{q(X)}{p(X)} \right) = \int_{\mathbb{R}^d} q(x) \ln(q(x)) dx - \int_{\mathbb{R}^d} q(x) \ln(p(x)) dx \quad (22)$$

La distance de Kullback-Leibler n'est pas une distance au sens mathématique du terme, mais elle permet d'établir un critère d'écart entre deux densités de probabilité. L'objectif de l'algorithme

d'optimisation est de minimiser la distance de Kullback-Leibler entre la densité auxiliaire optimale  $h_{opt}$  et la densité auxiliaire paramétrique  $h_\lambda$  :

$$D(h_{opt}, h_\lambda) = \int_{\mathbb{R}^d} \frac{\mathbf{1}_{\phi(x) > S} f(x)}{I_\phi} \ln\left(\frac{\mathbf{1}_{\phi(x) > S} f(x)}{I_\phi}\right) dx - \int_{\mathbb{R}^d} \frac{\mathbf{1}_{\phi(x) > S} f(x)}{I_\phi} \ln(h_\lambda(x)) dx \quad (23)$$

Cette équation peut s'écrire en terme d'espérance par :

$$D(h_{opt}, h_\lambda) = E_f \left[ \frac{\mathbf{1}_{\phi(X) > S}}{I_\phi} \ln\left(\frac{\mathbf{1}_{\phi(X) > S} f(X)}{I_\phi h_\lambda(X)}\right) \right] \quad (24)$$

La valeur optimale  $\lambda_{opt}$  qui minimise donc  $D(h_{opt}, h_\lambda)$  est définie par :

$$\lambda_{opt} = \underset{\lambda}{\operatorname{argmin}} \{D(h_{opt}, h_\lambda)\} \quad (25)$$

Cette équation est équivalente à trouver :

$$\lambda_{opt} = \underset{\lambda}{\operatorname{argmax}} \{E_f [\mathbf{1}_{\phi(X) > S} \ln(h_\lambda(X))]\} \quad (26)$$

Au lieu de maximiser directement,

$$\lambda \rightarrow E_f [\mathbf{1}_{\phi(X) > S} \ln(h_\lambda(X))] \quad (27)$$

on procède de manière itérative à partir d'une suite croissante

$$\gamma_0 < \gamma_1 < \gamma_2 < \dots < \gamma_t < \dots \leq S \quad (28)$$

de seuils (un choix adaptatif est donné par la suite). A l'itération  $t$ , où la valeur  $\lambda_{t-1}$  du paramètre est disponible, on demande juste que  $\phi(X)$  dépasse le seuil  $\gamma_t$ , et on maximise

$$\lambda \rightarrow E_f [\mathbf{1}_{\phi(X) > \gamma_t} \ln(h_\lambda(X))] = E_{h_{\lambda_{t-1}}} \left[ \mathbf{1}_{\phi(X) > \gamma_t} \frac{f(X)}{h_{\lambda_{t-1}}(X)} \ln(h_\lambda(X)) \right] \quad (29)$$

en utilisant la densité  $h_{\lambda_{t-1}}$  comme densité d'importance. En pratique, on maximise plutôt une version approchée obtenue par Monte Carlo et on définit ainsi

$$\lambda_t = \lambda_t^N = \underset{\lambda}{\operatorname{argmax}} \frac{1}{N} \sum_{i=1}^N \mathbf{1}_{\phi(X_i) > \gamma_t} \frac{f(X_i)}{h_{\lambda_{t-1}}(X_i)} \ln(h_\lambda(X_i)) \quad (30)$$

où les échantillons  $X_1, \dots, X_N$  sont distribués selon la densité  $h_{\lambda_{t-1}}$ . L'algorithme itératif suivant permet alors de déterminer la probabilité  $I_\phi$  par IS :

- 1) Poser  $h_{\lambda_0} = f$  et fixer  $t = 1$ .
- 2) Générer  $X_1, \dots, X_N$  selon la densité  $h_{\lambda_{t-1}}$ , appliquer la fonction  $\phi$  et poser  $Y_1 = \phi(X_1), \dots, Y_N = \phi(X_N)$ . Calculer le  $p$ -quantile de l'échantillon obtenu, c'est-à-dire ré-ordonner dans le sens croissant l'échantillon obtenu, de sorte que  $Y_{(1)} \leq \dots \leq Y_{(N)}$  et poser  $\gamma_t = \min(S, Y_{[pN]})$ .

$$3) \text{ Calculer } \lambda_t = \operatorname{argmax}_{\lambda} \left\{ \frac{1}{N} \sum_{i=1}^N \left[ \mathbf{1}_{\phi(X_i) > \gamma_t} \frac{f(X_i)}{h_{\lambda_{t-1}}(X_i)} \ln(h_{\lambda}(X_i)) \right] \right\}.$$

4) Si  $\gamma_t < S$ , il faut réitérer l'algorithme depuis l'étape 2) en posant  $t = t + 1$ .

5) Estimer la probabilité de l'évènement  $P(\phi(X) > S)$  par Importance Sampling à l'aide de l'équation suivante :

$$I_{\phi}^{CE} = \frac{1}{N} \sum_{i=1}^N \mathbf{1}_{\phi(X_i) > S} \frac{f(X_i)}{h_{\lambda_t}(X_i)} \quad (31)$$

La densité de probabilité  $h_{\lambda_t}$  déterminée par l'algorithme de CE minimise la distance de Kullback-Leibler avec la densité auxiliaire optimale de l'IS. Les performances d'estimation sont donc grandement améliorées par rapport à la densité  $f$ .

Appliquons la méthode de CE pour estimer  $I_5$  avec  $Y$  une variable aléatoire gaussienne centrée réduite. Nous considérons la famille  $h_{\lambda}$  des densités de probabilité gaussiennes. Le terme  $\lambda$  est donc un vecteur à 2 dimensions contenant la moyenne et la variance de la densité. Deux paramètres peuvent être notamment modifiés dans la méthode de CE :  $\rho$  qui représente la part d'échantillons utilisés dans la mise à jour des paramètres et  $N$  le nombre d'échantillons tirés à chaque itération. Plus  $\rho$  est élevé, plus la convergence est rapide, mais les paramètres  $\lambda$  ne sont pas estimés de manière précise. Il s'agit donc de faire un compromis. Nous avons choisi à titre d'exemple  $N = 1000$  et  $\rho = 0.9$ . Après 20 itérations de l'algorithme et donc 20000 échantillons utilisés, la figure 7 présente la densité auxiliaire optimisée par CE (gaussienne de moyenne 5.05 et d'écart-type 0.06) ainsi que la densité optimale. En utilisant 80000 échantillons de cette densité auxiliaire optimisée par CE et l'estimateur d'IS correspondant, on obtient l'estimée  $I_5^{CE} = 2.97 \cdot 10^{-7}$  à 0.03% d'écart-type relatif. Si l'on compare ce résultat à l'estimée par la méthode Monte Carlo, du tableau 2 pour  $N = 10^5$  échantillons, on montre effectivement que la méthode de CE est plus performante. Cette méthode est adaptée lorsque la probabilité recherchée est très rare ou bien lorsque l'on cherche une estimation très précise de la probabilité. En effet, la convergence de l'algorithme peut nécessiter un grand nombre d'échantillons. Cet algorithme est assez résistant à la dimensionalité du problème (nombre de paramètres des densités à estimer). Comme toute modélisation paramétrique, il est important que le modèle soit adapté au problème sinon l'optimisation est un échec.

### 4.3.3. Exponential twisting

La technique appelée exponential twisting, exponential tilting ou encore changement de mesure exponentielle [9, 27] est très connue en Importance Sampling et est notamment utilisée pour ses propriétés sur les sommes de variables aléatoires. Cette méthode permet ainsi de caractériser des dépassements de seuil par des marches aléatoires. Elle s'appuie sur la théorie des grandes déviations (borne de Chernoff) [4]. L'idée principale est de définir la densité auxiliaire  $h$  de la manière suivante :

$$h(x) = \exp(\theta \phi(x) - \mu(\theta)) f(x) \quad (32)$$



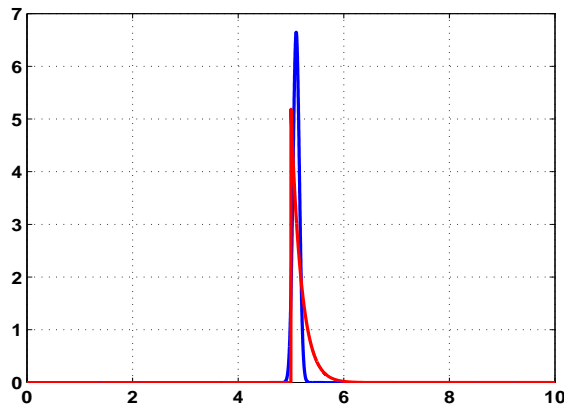


FIGURE 7. Densité auxiliaire optimale théorique (rouge) et densité auxiliaire optimisée par CE (bleu).

avec  $\mu(\theta) = \ln(M(\theta))$  et  $M(\theta) = E_f[\exp(\theta\phi(X))]$ . La variable de sortie  $Y = \phi(X)$  doit posséder des moments exponentiels, pour garantir que  $M(\theta)$  soit fini, au moins pour certaines valeurs de  $\theta$ . La densité  $h(x)$  varie en fonction du paramètre  $\theta$  et un paramètre optimal  $\theta_{opt}$  pour l'estimation de  $P(\phi(X) > S)$  peut alors être défini. Ce paramètre peut être estimé car celui-ci vérifie :

$$\mu'(\theta_{opt}) = S \quad (33)$$

Le paramètre  $\theta_{opt}$  est donc calculé numériquement.

Appliquons la méthode d'exponential twisting pour estimer  $I_5$  avec  $Y$  une variable aléatoire gaussienne centrée réduite. La figure 8 présente l'évolution des paramètres  $M(\theta)$  et  $\mu(\theta)$  en fonction de  $\theta$  avec  $N = 1000$  échantillons de la variable  $Y$ . Le paramètre  $\theta_{opt}$  est estimé numériquement à  $\hat{\theta} = 5.05$  à partir des échantillons. En appliquant la méthode d'IS, on obtient  $I_5^{ET} = 2.86 \cdot 10^{-7}$  avec une erreur de relative de 7.4%. Avec un faible coût d'optimisation, cette méthode permet de réduire fortement la variance d'estimation par rapport à un estimateur Monte Carlo pur.

#### 4.3.4. Importance Sampling adaptatif non paramétrique

L'objectif de l'algorithme Nonparametric Adaptive Importance Sampling (NAIS) est d'approcher la densité auxiliaire optimale de l'IS sans modélisation a priori [52, 32, 33]. Celle-ci est caractérisée à l'aide d'estimateurs à noyaux pondérés et permet donc une grande flexibilité par rapport à un modèle paramétrique. L'algorithme NAIS est donc le suivant :

- 1) Soit  $k = 0$  et  $g_0$  la densité initiale d'échantillonnage.
- 2) Dans un premier temps, on génère  $N_k$  échantillons  $X_1^{(k)}, \dots, X_{N_k}^{(k)}$  à partir de la PDF  $g_k$ .
- 3) On peut alors définir  $w_{k+1}(x) = \frac{\mathbf{1}_{\phi(x) > S} f(x)}{g_k(x)}$ .
- 4) On estime alors la probabilité  $I_{\phi,k}$  à l'itération  $k$  par

$$I_{\phi,k} = \frac{1}{m_{k+1}} \sum_{j=0}^k \sum_{i=1}^{N_j} w_{j+1}(X_i^{(j)}) \quad (34)$$

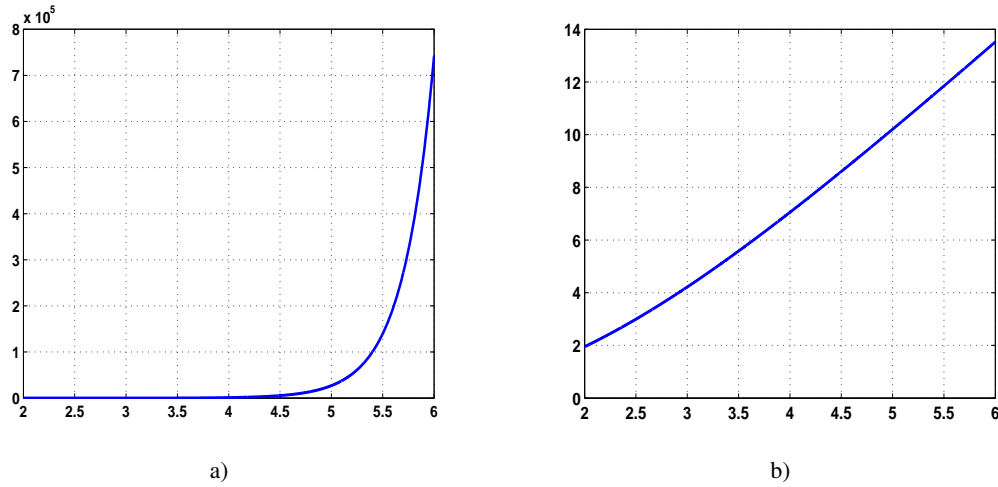


FIGURE 8. Evolution du paramètre  $M(\theta)$  a) et  $\mu(\theta)$  b) en fonction de  $\theta$  pour une densité gaussienne avec  $N = 1000$  échantillons.

en posant  $m_{k+1} = \sum_{j=0}^k N_j$ .

5) La PDF d'échantillonnage  $g_{k+1}$  est ensuite remise à jour de la manière suivante

$$g_{k+1}(x) = \frac{1}{m_{k+1} b_{k+1}^d I_{\phi,k}} \sum_{j=0}^k \sum_{i=1}^{N_j} w_{j+1}(X_i^{(j)}) K_d \left( \frac{X_i^{(j)} - x}{b_{k+1}} \right) \quad (35)$$

avec  $K_d$ , le noyau défini par une densité gaussienne centrée réduite de dimension  $d$ . Le terme  $b_{k+1}$  est une largeur de bande adaptée au noyau et optimisée par le critère AMISE (Asymptotic Mean Integrated Square Error) [48, 16].

6) Si  $k < k_{max}$ , retour à l'étape 2) de l'algorithme en posant  $k = k + 1$ .

7) On estime finalement la probabilité  $I_{\phi}^{NAIS}$  par

$$I_{\phi}^{NAIS} = \frac{1}{N_{k_{max}}} \sum_{i=1}^{N_{k_{max}}} \mathbf{1}_{\phi(X_i^{(k_{max})}) > S} \frac{f(X_i^{(k_{max})})}{g_{k_{max}}(X_i^{(k_{max})})} \quad (36)$$

L'utilisateur peut fixer le nombre désiré d'itérations en introduisant un nombre d'itérations maximum  $k_{max}$ . Considérons une nouvelle fois le cas simple où  $Y$  est une variable aléatoire gaussienne centrée réduite et estimons la probabilité  $I_5$ . Nous avons appliqué l'algorithme NAIS avec les paramètres suivants :

- $g_0 = N(0, 5)$  est une densité gaussienne de moyenne nulle et de variance 5
- Par mesure de simplicité, le nombre d'échantillons à chaque itération est fixé et égal à  $N_1 = N_2 = \dots = N_k = 500$
- $k_{max}$  est fixé à 9

Le tableau 8 présente l'estimation de  $I_5$  pour différentes itérations de l'algorithme NAIS. Après 4 itérations, la probabilité est en moyenne bien estimée et l'écart-type relatif décroît régulièrement. La convergence est obtenue pour  $k = 8$  où l'écart-type relatif est le plus faible. La figure 9 montre

itération $k$	$I_{5,k}^{NAIS}$	écart-type relatif
1	$2.66 \cdot 10^{-7}$	20.1%
2	$2.76 \cdot 10^{-7}$	11.3%
3	$2.79 \cdot 10^{-7}$	8.31%
4	$2.84 \cdot 10^{-7}$	2.02%
5	$2.85 \cdot 10^{-7}$	3.43%
6	$2.84 \cdot 10^{-7}$	2.74%
7	$2.84 \cdot 10^{-7}$	1.14%
8	$2.86 \cdot 10^{-7}$	1.06%
9	$2.85 \cdot 10^{-7}$	1.13%
10	$2.85 \cdot 10^{-7}$	1.05%

TABLEAU 8. Estimation moyenne répétée 100 fois de  $I_5^{NAIS}$  avec l'algorithme NAIS pour  $N_k = 500$  et  $k = 1 \dots 9$ .

l'évolution des densités  $g_k$  pour différentes valeurs de  $k$  ainsi que la densité optimale théorique  $h_{opt}$  (qui est disponible car  $I_\phi$  est connu théoriquement). Il y a peu de différences entre les densités obtenues par NAIS et la densité optimale, ce qui explique la bonne précision des estimations. Néanmoins, la précision de ces résultats est à relativiser car l'influence de la densité  $g_0$  peut être handicapante. Par exemple, appliquer l'algorithme NAIS avec  $g_0 = f$  s'avère en général inefficace, puisqu'on sait bien que

$$I_{\phi,1}^{NAIS} = \frac{1}{N_1} \sum_{i=1}^{N_1} \mathbf{1}_{\phi(X_i^{(0)}) > s} \quad (37)$$

risque d'être nul si  $X_1^{(1)} \dots X_N^{(1)}$  sont simulés à partir de la densité  $f$ . Dans un cas réaliste, il n'est donc pas aisé de déterminer une densité  $g_0$  appropriée.

#### 4.4. Application sur l'estimation d'une zone de retombée

Dans le cadre de l'estimation d'une zone de retombée, nous avons appliqué l'algorithme NAIS avec 4 itérations de 2500 simulations. Les résultats obtenus sont pertinents, avec un écart-type relatif très faible. Mais cela est dû au fait que la densité  $g_0$  initiale était adaptée. D'ailleurs, nous pouvons noter qu'avec cette même densité  $g_0$ , il n'est pas possible d'estimer la probabilité  $P(r > 4)$  car l'initialisation ne permet pas de générer des échantillons  $r_i$  tels  $r_i > 4$ . La dépendance à l'initialisation peut donc être très problématique dans cet algorithme. Pour s'affranchir de la densité  $g_0$ , une piste de recherche est proposée dans [33].

L'optimisation de la densité optimale de l'IS a aussi été effectuée par méthode CE dans les tableaux 10 et 11 en supposant un modèle de densité gaussien pour les entrées du modèle (optimisation de la moyenne et de la variance). Dans le cas du tableau 10, les résultats ne sont pas satisfaisants car le modèle de densité n'est pas adapté. En effet la queue positive et négative des entrées du modèle influent sur les événements rares. Or l'optimisation de la moyenne par CE biaise l'estimation vers la queue positive ou négative de la densité des entrées. Si au contraire, la moyenne est fixée et donc seule la variance est optimisée comme dans le cas du tableau 11, les résultats sont intéressants et améliorés fortement par rapport à une méthode de Monte Carlo classique. La contrainte du modèle aboutit à de moins bonnes performances d'estimation que l'algorithme NAIS comme le

seuil $S$ en km	$I_{\phi}^{NAIS}$	écart-type relatif
0.5	0.71	0.78%
1	0.28	2.1%
1.5	0.072	2.5%
2	0.0125	4.9%
2.5	0.0013	6.4%
3	$4.9 \cdot 10^{-5}$	12.1%
3.5	$1.5 \cdot 10^{-6}$	14.9%
4	?	?

TABLEAU 9. Estimation moyenne répétée 100 fois de  $I_{\phi}^{NAIS}$  avec  $N = 10000$  simulations par algorithme NAIS.

montre le tableau 9 mais est plus robuste pour l'estimation d'évènements très rares ( $P(r > 4)$ ). Si la dimension des entrées avait été plus importante, la CE aurait de même abouti à des résultats plus fiables car les algorithmes non paramétriques ne sont pas robustes à la grande dimension.

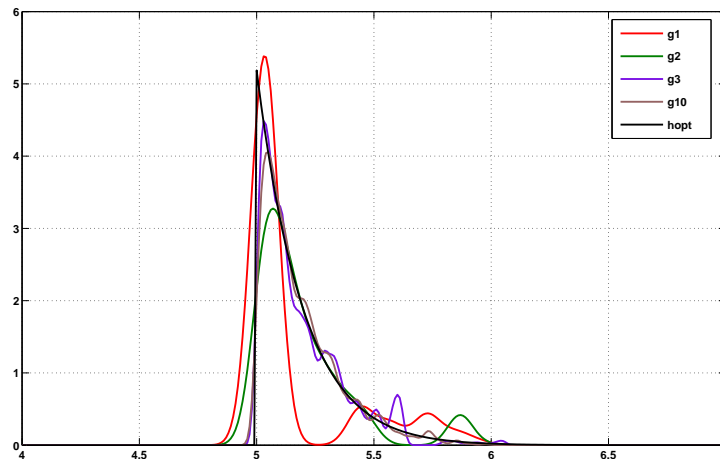


FIGURE 9. Densité optimale théorique  $h_{opt}$  (en noir) et les densités  $g_i$  par NAIS à l'aide de noyaux gaussiens pour  $i = \{1, 2, 3, 10\}$ , obtenues avec  $g_0 = N(0, 5)$ .

#### 4.5. Estimation de quantiles

Nous n'avons pas traité l'estimation de quantiles par IS. Celle-ci est notamment possible à l'aide d'algorithmes adaptatifs (optimisation CE) [19, 18]. Un estimateur de la fonction de répartition  $G_Y$  de la variable aléatoire  $Y = \phi(X)$  à l'aide d'IS est donné par l'équation suivante :

$$G_Y^{IS}(y) = \frac{1}{N} \sum_{i=1}^N \mathbf{1}_{\phi(X_i) \leq y} \frac{f(X_i)}{h(X_i)} \quad (38)$$

où  $h$  est la densité auxiliaire utilisée pour l'échantillonnage et  $f$  est la densité originale de la variable  $X$ . Il peut être intéressant en pratique de remplacer le dénominateur  $N$  par  $\sum_{i=1}^N \frac{f(X_i)}{h(X_i)}$  pour

seuil $S$ en km	$I_{\phi}^{CE}$	écart-type relatif
0.5	0.71	1.2%
1	0.28	1.5%
1.5	0.052	2.5%
2	0.0032	2.9%
2.5	0.0006	4.4%
3	$3.2 \cdot 10^{-6}$	7.1%
3.5	$0.7 \cdot 10^{-7}$	10.7%
4	$1.3 \cdot 10^{-9}$	20.2%

TABLEAU 10. Estimation moyenne répétée 100 fois de  $I_{\phi}^{CE}$  avec  $N = 50000$  simulations par IS avec optimisation CE sur la moyenne et la variance de densités gaussiennes.

seuil $S$ en km	$I_{\phi}^{CE}$	écart-type relatif
0.5	0.72	0.5%
1	0.29	1.3%
1.5	0.075	2.6%
2	0.0130	3.8%
2.5	0.0012	7.2%
3	$4.7 \cdot 10^{-5}$	22.4%
3.5	$1.3 \cdot 10^{-6}$	25.6%
4	$8.4 \cdot 10^{-8}$	?

TABLEAU 11. Estimation moyenne répétée 100 fois de  $I_{\phi}^{CE}$  avec  $N = 50000$  simulations par IS avec optimisation CE sur la variance de densités gaussiennes.

forcer la fonction de répartition  $G_Y^{IS}(y) \rightarrow 1$  quand  $y \rightarrow \infty$ . Cet estimateur est non biaisé dès que le support de la densité  $f$  est inclus dans le support de  $h$ . La variance de cet estimateur est donnée par l'équation suivante :

$$\text{Var}_h(G_Y^{IS}(y)) = \frac{1}{N} (E_f(\mathbf{1}_{\phi(X) \leq y} w(X)^2) - G_Y^2(y)) \quad (39)$$

où  $w(x) = \frac{f(x)}{h(x)}$ . La densité optimale permettant d'obtenir une variance minimale s'obtient de la même manière que dans le cas de probabilités.

$$h_{opt}(x) = \frac{\mathbf{1}_{\phi(x) \leq y} f(x)}{G_Y(y)} \quad (40)$$

La densité optimale  $h_{opt}$  n'est bien sûr pas utilisable, mais des procédures adaptatives permettent d'optimiser la variance d'estimation. La stratégie consiste donc à déterminer une PDF qui soit efficace pour l'estimation du  $\alpha$ -quantile  $q_{\alpha}$  :

$$P(\phi(X) \leq q_{\alpha}) = \int \mathbf{1}_{\phi(z) \leq q_{\alpha}} f(z) dz = \alpha \quad (41)$$

#### 4.6. Conclusion

La méthode d'Importance Sampling est simple à mettre en place pour estimer des événements rares à l'aide d'une densité auxiliaire. La principale difficulté de l'algorithme consiste justement à

estimer cette densité à l'aide des méthodes d'optimisation que nous avons présentées. Le nombre de simulations nécessaires pour l'optimisation peut être élevé, mais il est faible vis-à-vis de la probabilité estimée. Dans le cas général, la dimensionnalité est souvent problématique car il est difficile d'échantillonner un espace à grande dimension. En pratique, lorsque le rééchantillonnage des entrées par l'utilisateur est possible mais que ce nombre de simulations est tout de même limité, la méthode d'Importance Sampling est tout à fait adaptée.

## 5. Importance Splitting

Le Splitting [17, 25, 26, 7, 6, 35] est une méthode d'estimation d'espérances mathématiques tout particulièrement utile dans le cadre des événements rares et des densités de probabilités conditionnelles. L'idée fondamentale est de diviser, *to split* en anglais, la difficulté du problème. Plutôt que de s'attaquer directement au problème complexe à résoudre, on le réexprime en une suite de problèmes plus simples, de manière à ce que l'enchaînement de leurs solutions permette de résoudre le problème originel.

### 5.1. Principe

Dans le cas qui nous intéresse, on cherche à estimer

$$I_\phi = P[\phi(X) > S] = P[X \in \mathbf{A}]$$

où  $\mathbf{A} = \{x \in \mathbb{R}^d | \phi(x) \geq S\}$ . Pour effectuer la reformulation de la probabilité voulue, la technique d'importance splitting repose sur le choix d'une suite de sous-domaines emboîtés de  $\mathbb{R}^d$  ayant  $\mathbf{A}$  pour plus petit élément, soit

$$\mathbb{R}^d = \mathbf{A}_0 \supset \mathbf{A}_1 \supset \dots \supset \mathbf{A}_{n-1} \supset \mathbf{A}_n = \mathbf{A}$$

définis par

$$\mathbf{A}_k = \{x \in \mathbb{R}^d | \phi(x) \geq S_k\}$$

pour tout  $k = 1, \dots, n$  via une suite croissante de seuils

$$S_1 \leq \dots \leq S_{n-1} \leq S_n = S$$

On réécrit alors  $I_\phi$ , la probabilité cible, à l'aide de la formule de Bayes :

$$I_\phi = \prod_{k=1}^n P(X \in \mathbf{A}_k | X \in \mathbf{A}_{k-1})$$

En parallèle de l'estimation des probabilités conditionnelles, on approche les densités conditionnelles de la variable  $X$ , conditionnée à appartenir à  $\mathbf{A}_k$ , définies par

$$f_k(x) = \frac{\mathbf{1}_{\phi(x) \geq S_k} f(x)}{P(\phi(X) \geq S_k)}$$

### 5.2. Estimation des ensembles $\mathbf{A}_k$

Sous sa forme d'origine, il faut définir a priori, dans le splitting, les seuils  $S_k$  à franchir, ce qui requiert une connaissance de  $\phi$  souvent indisponible. Les ensembles  $\mathbf{A}_k$  sont donc déterminés de manière adaptative à l'aide de quantiles. Etant donné un échantillon de points  $X_1^{(k-1)}, \dots, X_R^{(k-1)}$  distribués selon  $f_{k-1}$ , le seuil suivant  $S_k$  est défini comme le quantile empirique de niveau arbitraire  $\beta_k$  de  $\phi(X_1^{(k-1)}), \dots, \phi(X_R^{(k-1)})$ , ce qui permet de poser

$$\mathbb{P}[X \in \mathbf{A}_k | X \in \mathbf{A}_{k-1}] = 1 - \beta_k$$

Cela assure que tous les seuils seront franchis et donc que le critère d'arrêt de l'algorithme sera atteint. L'optimalité d'une telle stratégie a été étudiée dans [22] où il est montré qu'il faut choisir  $\beta_k = \beta$  pour tout  $k$ , pour minimiser la variance de l'estimation. Le niveau  $\beta$  apparaît alors comme un moyen de contrôle du compromis entre coût de simulation et variance. La valeur de  $\beta$  est fixée dans l'article à 0.2.

### 5.3. Génération des échantillons selon les densités conditionnelles $f_k$

La simulation selon la densité  $f_k$ , inconnue a priori, est effectuée à l'aide d'un noyau markovien  $f$ -réversible [50, 6]. Un noyau markovien  $f$ -réversible est donné par une densité de transition  $M$  telle que

$$f(x)M(x, \xi) = f(\xi)M(\xi, x) \text{ pour tout } x, \xi \in \mathbb{R}^d$$

Si la variable  $X$  a pour densité  $f_k$ , alors on peut montrer que la variable  $X'$  construite selon le mécanisme de proposition-sélection ci-dessous a également pour densité  $f_k$  : on génère une variable  $\Xi$  selon la densité de transition  $M(X, \cdot)$ , et on pose

$$X' = \begin{cases} \Xi, & \text{si } \phi(\Xi) \geq S_k \\ X, & \text{sinon} \end{cases} \quad (42)$$

c'est-à-dire qu'on accepte la proposition  $\Xi$  si elle appartient à  $\mathbf{A}_k$  et on conserve la valeur initiale  $X$  sinon. Un noyau markovien  $f$ -réversible permet donc de générer un échantillon de variables distribuées selon  $f_k$ , à partir d'un échantillon de variables déjà distribuées selon  $f_k$ , avec l'avantage que l'échantillon transformé est en principe moins dépendant que l'échantillon initial. Itérer ce mécanisme de proposition-sélection est un moyen souvent efficace de réduire significativement la dépendance au sein d'un échantillon distribué selon  $f_k$ , et donc de réduire la variance de l'estimateur.

Dans le cas où  $X$  est une variable aléatoire gaussienne centrée réduite de dimension  $d$ , un noyau markovien réversible possible est donné par la densité de transition gaussienne  $M(x, \cdot)$ , de moyenne  $\frac{x}{\sqrt{1+c^2}}$  et de matrice de covariance  $\frac{c^2}{1+c^2}I_d$ , et générer une variable aléatoire  $\Xi$  selon la densité  $M(X, \cdot)$  revient simplement à générer une variable aléatoire gaussienne centrée réduite  $W$ , et à poser

$$\Xi = \frac{X + cW}{\sqrt{1+c^2}} \quad (43)$$

Le choix d'une valeur appropriée du noyau  $c$  n'est pas évident et est toujours une question ouverte. La valeur de  $c$  peut aussi être déterminée de manière adaptative. Dans cet article, la valeur de  $c$  est fixée à 0.5.

#### 5.4. Algorithme

Il s'agit donc en pratique d'exécuter le programme suivant pour estimer la probabilité  $P(\phi(X) > S) = P(X \in \mathbf{A})$  :

- 1) On pose  $k = 0$ .
- 2) On génère  $R$  échantillons iid  $X_1^{(0)}, \dots, X_R^{(0)}$  selon la densité  $f$ .
- 3) On définit  $S_1$  comme le  $\beta$ -quantile empirique de  $\phi(X_1^{(0)}), \dots, \phi(X_R^{(0)})$ .
- 4) Tant que  $S_{k+1} < S$ , on procède de la manière suivante :
  - a) on pose  $k = k + 1$  et on remplace les échantillons dont l'image par  $\phi$  est inférieure au seuil  $S_k$  par des échantillons choisis, par tirage uniforme avec remise, parmi les échantillons dont l'image par  $\phi$  est supérieure au seuil  $S_k$ .
  - b) On applique le mécanisme de proposition-sélection de l'équation (42) autant de fois que l'on souhaite, à tous les échantillons pour obtenir la population  $X_1^{(k)}, \dots, X_R^{(k)}$ .
  - c) On définit  $S_{k+1}$  comme le  $\beta$ -quantile empirique de  $\phi(X_1^{(k)}), \dots, \phi(X_R^{(k)})$ .
- 7) On estime alors la probabilité avec :

$$I_{\phi}^{ISP} = (1 - \beta)^k \times \left( \frac{1}{R} \sum_{i=1}^R \mathbf{1}_{\phi(X_i^{(k)}) > S} \right) \quad (44)$$

Pour l'estimation de quantiles, l'algorithme est similaire et seuls les critères de sortie de boucle et l'estimation finale changent.

#### 5.5. Application sur un cas simple

Considérons le cas simple où  $Y$  est une variable aléatoire gaussienne centrée réduite et appliquons l'algorithme d'Importance Splitting. Le tableau 12 présente l'estimation des probabilités de passage de seuil de la variable aléatoire  $Y$ . Les résultats sont améliorés par rapport à une méthode de Monte Carlo classique mais peuvent paraître décevants par rapport à l'algorithme NAIS. Néanmoins, il faut relativiser très fortement cet écart de performances, car aucune hypothèse a priori sur la densité d'échantillonnage n'a été effectuée pour l'algorithme d'Importance Splitting, contrairement au cas du NAIS. Dans un cadre réaliste, l'algorithme d'Importance Splitting peut donc être bien plus performant que l'algorithme NAIS. L'estimation de quantiles est aussi possible de manière rapide grâce à un bon positionnement de la suite des espaces  $\mathbf{A}_k$ . La méthode de seuillage adaptatif permet de résoudre ce problème et on peut ainsi estimer les quantiles désirés. Le tableau 13 présente l'estimation du quantile à  $1 - 10^{-7}$  avec une forte précision.

#### 5.6. Application sur l'estimation d'une zone de retombée

Dans le cas de l'estimation de zone de retombée d'engin, nous avons obtenu les résultats du tableau 14. La méthode d'Importance Splitting permet d'estimer des quantiles rares qu'il n'était pas possible d'évaluer avec les autres méthodes considérées dans cet article. La précision d'estimation semble être plus faible que dans le cas de l'algorithme NAIS, mais il faut noter



N	$I_4^{ISP}$	$I_5^{ISP}$
$10^4$	$2.39 \cdot 10^{-5}$ (43%)	$1.38 \cdot 10^{-7}$ (67%)
$10^5$	$3.00 \cdot 10^{-5}$ (13%)	$2.39 \cdot 10^{-7}$ (22%)
$10^6$	$3.16 \cdot 10^{-5}$ (4.4%)	$2.81 \cdot 10^{-7}$ (5.2%)
$10^7$	$3.16 \cdot 10^{-5}$ (1.4%)	$2.86 \cdot 10^{-7}$ (1.8%)
théorie	$3.17 \cdot 10^{-5}$	$2.86 \cdot 10^{-7}$

TABLEAU 12. Estimation moyenne répétée 100 fois de  $I_S$  et des erreurs relatives d'estimation par Importance Splitting pour différentes valeurs de  $N$  et de  $S$  pour une variable gaussienne centrée réduite.

N	$q_{0.9999999}$
$10^4$	5.04 (2.2%)
$10^5$	5.18 (0.8%)
$10^6$	5.19 (0.21%)
$10^7$	5.19 ( $7.4 \cdot 10^{-2}\%$ )
théorie	5.20

TABLEAU 13. Estimation moyenne répétée 100 fois de quantiles rares d'une variable gaussienne centrée réduite par Importance Splitting pour différentes valeurs de  $N$ .

que l'Importance Splitting ne considère pas une densité d'initialisation ce qui est un atout non négligeable. De nombreux échantillons sont nécessaires pour permettre une bonne convergence de la méthode. Il est donc intéressant de la mettre en place lorsque le nombre d'échantillons disponible n'est pas limité.

### 5.7. Conclusion sur l'Importance Splitting

L'Importance Splitting est un algorithme performant pour l'estimation de probabilités faibles. Néanmoins, sa complexité nécessite la génération de nombreux échantillons avant la convergence. Cette méthode ne fait aucune hypothèse d'initialisation et peut donc être utilisée dans un cadre très général sans connaissance de la fonction  $\phi$  et sans hypothèse sur la densité optimale d'échantillonnage. Mais cette méthode contient de nombreux paramètres à régler et une expertise est donc nécessaire pour obtenir des résultats optimaux.

seuil $S$ en km	$I_\phi^{ISP}$	écart-type relatif
0.5	0.72	1.3%
1	0.28	1.9%
1.5	0.074	3.5%
2	0.0123	4.2%
2.5	0.0012	6.1%
3	$5.1 \cdot 10^{-5}$	8.2%
3.5	$1.3 \cdot 10^{-6}$	12.1%
4	$8.4 \cdot 10^{-8}$	15%

TABLEAU 14. Estimation moyenne répétée 100 fois de  $I_\phi^{ISP}$  avec  $N = 100000$  simulations par Importance Splitting.

Méthode	Rééchantillonnage nécessaire	Quantiles et probabilités estimés	Caractéristiques
Théorie des valeurs extrêmes	NON	tous les quantiles ou probabilités rares peuvent être estimés	simplicité de mise en œuvre mais résultats à valider et à analyser.
Importance Sampling	OUI	la PDF auxiliaire $h$ est à adapter pour chaque probabilité et quantile estimé	optimisation nécessaire pour obtenir de bons résultats et sensible à la dimensionnalité
Importance Splitting	OUI	plusieurs quantiles ou probabilités peuvent être estimés en une fois	expertise nécessaire pour définir des paramètres de simulation efficaces

TABLEAU 15. Résumé des domaines d'emploi des différentes méthodes d'estimation de probabilités et de quantiles rares.

seuil $S$ en km	$I_{\phi}^{NAIS}$	écart-type relatif
0.5	0.71	0.78%
1	0.28	2.1%
1.5	0.072	2.5%
2	0.0125	4.9%
2.5	0.0013	6.4%
3	$4.9 \cdot 10^{-5}$	12.1%
3.5	$1.5 \cdot 10^{-6}$	14.9%
4	?	?

$N = 10000$  avec l'algorithme NAIS

seuil $S$ en km	$I_{\phi}^{MC}$	écart-type relatif
0.5	0.72	0.64%
1	0.29	1.4%
1.5	0.074	3.2%
2	0.0125	8.1%
2.5	0.0013	26.1%
3	$4.6 \cdot 10^{-5}$	132.6%
3.5	$1.0 \cdot 10^{-6}$	1000%
4	?	?

$N = 10000$  avec la méthode Monte Carlo

seuil $S$ en km	$I_{\phi}^{ISP}$	écart-type relatif
0.5	0.72	1.3%
1	0.28	1.9%
1.5	0.074	3.5%
2	0.0123	4.2%
2.5	0.0012	6.1%
3	$5.1 \cdot 10^{-5}$	8.2%
3.5	$1.3 \cdot 10^{-6}$	12.1%
4	$8.4 \cdot 10^{-8}$	15%

$N = 100000$  avec importance splitting

seuil $S$ en km	$I_{\phi}^{IVE}$
0.5	?
1	?
1.5	0.070
2	0.012
2.5	0.0016
3	$7.8 \cdot 10^{-5}$
3.5	$4.2 \cdot 10^{-6}$
4	$1.3 \cdot 10^{-7}$

$N = 10000$  avec la théorie des valeurs extrêmes

TABLEAU 16. Récapitulatif des résultats obtenus dans le cas de la zone de retombée.

## 6. Conclusion

Nous avons étudié dans cet article les principales méthodes permettant d'estimer des probabilités et des quantiles rares. Celles-ci se déclinent donc en trois classes principales : la théorie des valeurs extrêmes, l'Importance Sampling et l'Importance Splitting. Leur domaine d'emploi est en somme assez complémentaire sur le plan pratique. Le tableau 15 fait la synthèse des informations nécessaires pour chaque méthode. Celles-ci ont été testées dans le cadre de l'estimation de zones de retombée et leurs résultats sont rappelés dans le tableau 16.

Une analyse de la fonction  $\phi$  permet aussi de générer des événements rares à moindre coût de calcul. On peut ainsi chercher à caractériser un modèle approché  $\hat{\phi}$  à l'aide de krigeage ou de réseaux de neurones, ou bien à déterminer des paramètres influents (indices de Sobol, plan d'expériences...) qui pourraient donc améliorer la précision et la convergence des algorithmes d'événements rares. Les méthodes statistiques avancées pour l'estimation d'un événement rare à l'aide d'un code de calcul (vu comme une boîte noire) sont en cours de développement et n'ont pas été traités dans cet article [37, 47, 41, 5, 51, 38]. Parmi ces travaux, beaucoup sont consacrés à l'exploitation d'un métamodèle. Contrairement aux méthodes décrites dans ce papier, leur intérêt est de chercher, préalablement à l'estimation de l'événement rare ou en parallèle, à modéliser la relation entre la sortie  $Y$  et les entrées  $X$ , puis à l'exploiter dans une phase de planification adaptative.

## 7. Remerciements

Ces travaux ont été en partie financés par le Ministère de l'Economie, de l'Industrie et de l'Emploi dans le cadre du projet CSDL (Complex System Design Lab) du Pôle de Compétitivité System@tic - Paris Région. Les auteurs remercient aussi les deux rapporteurs anonymes pour leurs commentaires.

## Références

- [1] I. BEICHL et F. SULLIVAN : The importance of importance sampling. *Computing in science and engineering*, 1:pp.71–73, Mars/Avril 1999.
- [2] P. H. BORCHERDS : Importance sampling : An illustrative introduction. *European Journal of Physics*, pages 405–411, April 2000.
- [3] Z. I BOTEV, D.P. KROESE et T. TAIMRE : Generalized cross-entropy methods with applications to rare-event simulation and optimization. *Simulation*, 11:785 – 806, 2007.
- [4] J. A. BUCKLEW : *Large Deviation Techniques in Decision, Simulation, and Estimation*. John Wiley, New York, 1990.
- [5] C. CANNAMELA, J. GARNIER et B. IOOSS : Controlled stratification for quantile estimation. *Annals of Applied Stats*, 2:1554–1580, 2008.
- [6] F. CEROU, P. DEL MORAL, T. FURON et A. GUYADER : Sequential Monte Carlo for rare event estimation. *Statistics and Computing*, 2011.
- [7] F. CEROU et A. GUYADER : Adaptive particle techniques and rare event estimation. volume 19, pages 65–72. ESAIM : Proceedings, 2007.
- [8] S.G. COLES : *An introduction to statistical modeling of extreme values*. Springer, New York, 2001.
- [9] H. E. DANIELS : Saddlepoint approximations in statistic. *Ann. Math. Statist*, 5:631–650, 1954.
- [10] A.C. DAVISON et R.L. SMITH : Models for exceedances over high thresholds (with discussion). *Journal of the Royal Statistical Society*, 52:393–442, 1990.

- [11] L. de HAAN et J. PICKANDS : Stationary min-stable stochastic processes. *Probability Theory and Related Fields*, 7:477–492, 1986.
- [12] M. DENNY : Introduction to importance sampling in rare-event simulations. *European Journal of Physics*, pages 403–411, July 2001.
- [13] P. EMBRECHTS, C. KLUPPELBERG et T. MIKOSCH : *Modelling Extremal Events for Insurance and Finance*. Springer Verlag, Berlin, 1997.
- [14] R.A. FISHER et L.H.C. TIPPETT : On the estimation of the frequency distributions of the largest or smallest member of a sample. *Proceedings of the Cambridge Philosophical Society*, 1928.
- [15] M. GILLI et E. KËLLEZI : An application of extreme value theory for measuring risk. *Computational Economics*, 27:207–228, 2006.
- [16] I. K. GLAD, N. L. HJORT et N. G. USHAKOV : Mean-squared error of kernel estimators for finite values of the sample size. *Journal of Mathematical Sciences*, 146:5977–5983, 2007.
- [17] P. GLASSERMAN, P. HEIDELBERGER, P. SHAHABUDDIN et T. ZAJIC : Splitting for rare event simulation : analysis of simple cases. *In Proceeding of the 1996 Winter Simulation Conference*, pages 302–308, 1996.
- [18] P. GLYNN : Importance sampling for Monte Carlo estimation of quantiles. *In Proceedings of the Second International Workshop on Mathematical Methods in Stochastic Simulation and Experimental Design, Saint Petersburg*, pages 180–185. Publishing House of Saint Petersburg University, 1996.
- [19] T.C. HESTERBERG et B.L. NELSON : Control variates for probability and quantile estimation. *Management Science*, 44(9):1295–1312, 1998.
- [20] T. HOMEM-DE-MELLO et R. Y. RUBINSTEIN : Estimation of rare event probabilities using cross-entropy. *In Proceedings of the 2002 Winter Simulation Conference, San Diego*, pages 310–319, 2002.
- [21] J.R.M. HOSKING et J.R. WALLIS : Parameter and quantile estimation for the Generalized Pareto Distribution. *Technometrics*, 29:339–349, 1987.
- [22] A. LAGNOUX : Rare event simulation. *Probability in the Engineering and Informational science*, 20:45–66, 2006.
- [23] M.R. LEADBETTER, G. LINDGREN et H. ROOTZËN : *Extremes and related properties of random sequences and series*. Springer Verlag, New York, USA, 1983.
- [24] M.R. LEADBETTER et H. ROOTZËN : Extremal theory for stochastic processes. *Annals of Probability*, 16:431–478, 1988.
- [25] P. L'ËCUYER, V. DEMERS et B. TUFFIN : Splitting for rare event simulation. *In Proceeding of the 2006 Winter Simulation Conference, Monterey*, pages 137–148, 2006.
- [26] P. L'ËCUYER, F. LE GLAND, P. LEZAUD et B. TUFFIN : Splitting methods. *In Gerardo RUBINO et Bruno TUFFIN, éditeurs : Monte Carlo Methods for Rare Event Analysis*, chapitre 3, pages 39–61. John Wiley & Sons, Chichester, 2009.
- [27] R. LUGANNANI et S. RICE : Saddlepoint approximation for the distribution of the sum of independent random variables. *Adv. in Appl. Probab*, 12:475–490, 1980.
- [28] A. MCNEIL et T. SALADIN : The peaks over threshold method for estimating high quantiles of loss distributions. *Proceedings of the 28th International ASTIN Colloquium, Cairns*, 1997.
- [29] G. A. MIKHAILOV : *Parametric Estimates by the Monte Carlo Method*. VSP, Utrecht (NED), 1999.
- [30] A. MILLET : *Méthode de Monte Carlo*. Université Paris VI et Paris VII, Polycopié, 2003.
- [31] I. S. MOLCHANOV : Empirical estimation of distribution quantiles of random closed sets. *Theory Proba. Appli.*, 35:594–600, 1990.
- [32] J. MORIO : How to approach the importance sampling density. *Eur. J. Phys*, 31:41–48, 2010.
- [33] J. MORIO : Non parametric adaptive importance sampling of the probability estimation of launcher impact position. *Reliability Engineering and System Safety*, 96(1):178–183, 2011.
- [34] J. MORIO et R. PASTEL : Sampling technique for launcher impact safety zone estimation. *Acta Astronautica*, 66(5-6):736–741, 2010.
- [35] J. MORIO, R. PASTEL et F. LE GLAND : An overview of importance splitting for rare event simulation. *European Journal of Physics*, 31:1295–1303, 2010.
- [36] H. NIEDERREITER et J. SPANIER : *Monte Carlo and Quasi-Monte Carlo Methods*. Springer, 2000.
- [37] J. OAKLEY : Estimating percentiles of uncertain computer code outputs. *Appl. Statist.*, 53:83–93, 2004.

- [38] V. PICHENY, D. GINSBOURGER, O. ROUSTANT et R.T. HAFTKA : Adaptive designs of experiments for accurate approximation of a target region. *J. Mech. Des.*, 137(7), 2010.
- [39] J. PICKANDS : Statistical inference using extreme order statistics. *Annals of Statistics*, 3:119–131, 1975.
- [40] M. PIERA MARTINEZ, E. VAZQUEZ, E. WALTER, G. FLEURY et R. KIELBASA : Estimation of extreme values, with application to uncertain systems. In *Proceedings of the 14th IFAC Symposium on System Identification (SYSID)*, Newcastle, pages 1027–1032. IFAC, 2006.
- [41] P. RANJAN, D. BINGHAM et G. MICHAILIDIS : Sequential experiment design for contour estimation from complex computer codes. *Technometrics*, 50:527–541, 2008.
- [42] C. P. ROBERT et G. CASELLA : *Monte Carlo Statistical Methods*. Springer, New York, 2005.
- [43] R.Y. RUBINSTEIN : *Simulation and the Monte Carlo Method*. John Wiley, 1981.
- [44] R.Y. RUBINSTEIN : Optimization of computer simulation models with rare events. *European Journal of Operations Research*, 99:89–112, 1997.
- [45] R.Y. RUBINSTEIN : The cross-entropy method for combinatorial and continuous optimization. *Methodology and Computing in Applied Probability*, 2:127–129, 1999.
- [46] R.Y. RUBINSTEIN et D.P. KROESE : *The Cross-Entropy Method : A Unified Approach to Combinatorial Optimization, Monte-Carlo Simulation and Machine Learning*. Springer Verlag, New York, 2004.
- [47] B. RUTHERFORD : A response-modeling alternative to surrogate models for support in computational analyses. *Reliability Engineering and System Safety*, 91:1322–1330, 2006.
- [48] B.W. SILVERMAN : *Density Estimation for Statistics and Data Analysis*. London : Chapman and Hall, 1986.
- [49] I. M. SOBOL : *A Primer for the Monte Carlo Method*. CRC Press, Boca Raton, Fl., 1994.
- [50] L. TIERNEY : Markov chains for exploring posterior distributions. *Annals of Statistics*, 22:1701–1762, 1994.
- [51] E. VAZQUEZ et J. BECT : A sequential Bayesian algorithm to estimate a probability of failure. In *Proceedings of the 15th IFAC Symposium on System Identification (SYSID)*, Saint-Malo, pages 546–550. IFAC, 2009.
- [52] P. ZHANG : Nonparametric importance sampling. *Journal of the American Statistical Association*, 91(434):1245–1253, septembre 1996.