Estimation de la probabilité d'événements rares

Yves AUFFRAY[1] , Pierre BARBILLON[2] , Jean-Michel MARIN[3]

Université Paris-Sud 11
Projet SELECT (Inria Saclay -île-de-France)

pierre.barbillon@math.u-psud.fr

[1]Dassault Aviation, Université Paris-sud 11
[2]Université Paris-Sud 11, INRIA Saclay
[3]Université Montpellier 2

Problématique

Soient

- $\mathbf{X}$ est une variable aléatoire à valeurs dans $E \subset \mathbb{R}^d$ de loi connue, simulable
- g sa densité par rapport à la mesure de Lebesgue.
- $f : E \subset \mathbb{R}^d \to \mathbb{R}_+$ une fonction "boîte noire" coûteuse
- $R = \{\mathbf{x} \in E \subset \mathbb{R}^d : f(\mathbf{x}) \leq \rho\}$ où $\rho > 0$ est fixé.

$\{\mathbf{X} \in R\}$ est un événement redouté et rare.

Objectif : Borner supérieurement $t_\rho = \mathbb{P}(\mathbf{X} \in R) = \mathbb{P}(f(\mathbf{X}) < \rho)$ avec un bon niveau de confiance.
Compte tenu de son coût d'évaluation, on dispose d'un "budget"maximal de N appels à f.

Limite de l'estimateur de Monte-Carlo

Pour un N-échantillon $(\mathbf{X}_1, \ldots, \mathbf{X}_N)$ de copies de $\mathbf{X}$, l'estimateur de Monte-Carlo est

$$\hat{t}_\rho^{MC} = 1/N \sum_{i=1}^{N} \mathbb{I}_R(\mathbf{X}_i)\,.$$

$$\begin{aligned} \mathbb{E}(\hat{t}_\rho^{MC}) &= \mathbb{P}(X \in R) = t_\rho, \\ \mathrm{Var}(\hat{t}_\rho^{MC}) &= \frac{1}{N} t_\rho (1 - t_\rho). \end{aligned}$$

Si $t_\rho << 1/N$, $\hat{t}_\rho^{MC}$ sera nul, très probablement et l'intervalle de confiance associé à l'estimateur, de forme $[0, t_\rho^+]$ pourra ne pas être assez précis.

Exemple : si $t_\rho = 4.7 \cdot 10^{-4}$ et $N = 100$,

- $\hat{t}_\rho^{MC} = 0$ avec probabilité $(1 - t_\rho)^N = 0.95$,
- le cas échéant, l'intervalle de confiance à 99% est $[0; 1 - (1/100)^{1/N}] = [0; 0.045]$.

Plan

Plan

Modèle

La fonction f est considérée comme la réalisation d'un processus gaussien $(Y_{\mathbf{x}})_{x \in E}$:

$$Y_{\mathbf{x}} = \sum_{i=1}^{p} \beta_i h_i(\mathbf{x}) + Z(\mathbf{x}) = H(\mathbf{x})^T \boldsymbol{\beta} + Z(\mathbf{x}),$$

avec

- h_i des fonctions de régression choisies,
- β_i des paramètres réels,
- Z un processus gaussien centré caractérisé par sa fonction de covariance $\text{cov}(Z(\mathbf{x}), Z(\mathbf{x}')) = \sigma^2 K(\mathbf{x}, \mathbf{x}')$ où K est un noyau symétrique positif tel que pour tout $\mathbf{x}$, $K(\mathbf{x}, \mathbf{x}) = 1$.

Loi conditionnelle

On dispose des évaluations $\{y_1 = f(\mathbf{x}_1), \ldots, y_n = f(\mathbf{x}_n)\}$ aux points du plan d'expérience fixé $D = \{\mathbf{x}_1, \ldots, \mathbf{x}_n\} \subset E^n$.

Processus a posteriori

On note $(Y_{\mathbf{x}}^D)_{\mathbf{x}\in E}$ le processus gaussien correspondant au processus gaussien a priori $(Y_{\mathbf{x}})_{\mathbf{x}\in E}$ conditionné à l'observation de $y_{\mathbf{x}_1}, \ldots, y_{\mathbf{x}_n}$.

- $\forall \mathbf{x},\ \mathbb{E}(Y_{\mathbf{x}}^D) = \hat{f}(\mathbf{x}; y_{\mathbf{x}_1}, \ldots, y_{\mathbf{x}_n}) = H(\mathbf{x})^T\beta + \Sigma_{\mathbf{x}D}^T\Sigma_{DD}^{-1}(Y_D - H_D\beta)$,
- $\forall \mathbf{x}, \mathbf{x}',\ \mathrm{cov}(Y_{\mathbf{x}}^D, Y_{\mathbf{x}'}^D) = \sigma^2\left(K(\mathbf{x}, \mathbf{x}') - \Sigma_{\mathbf{x}D}^T\Sigma_{DD}^{-1}\Sigma_{\mathbf{x}'D}\right)$,

où $H_D = (H(\mathbf{x}_1), \ldots H(\mathbf{x}_n))^T$, $(\Sigma_{DD})_{1\leq i,j\leq n} = K(\mathbf{x}_i, \mathbf{x}_j) = \mathrm{corr}(Y_{\mathbf{x}_i}, Y_{\mathbf{x}_j})_{1\leq i,j\leq n}$ et $\Sigma_{\mathbf{x}D} = (\mathrm{corr}(Y_{\mathbf{x}_1}, Y_{\mathbf{x}}))_{1\leq i\leq n}^T$.

Remarque

$$\forall \mathbf{x},\ \ \mathrm{cov}(Y_{\mathbf{x}}^D, Y_{\mathbf{x}}^D) = \mathrm{MSE}(\mathbf{x})$$

Validation

Jones et al. (1998) proposent une méthode de validation croisée loo afin de valider le modèle réduit.

On note $\hat{f}_{-i}$ l'approximation de f par krigeage construite sur $\{\mathbf{x}_1, \ldots, \mathbf{x}_{i-1}, \mathbf{x}_{i+1}, \ldots, \mathbf{x}_n\}$ et $s_{-i}(.) = \sqrt{\text{MSE}_{-i}(.)}$ la fonction racine carrée des erreurs quadratiques de prédiction correspondante.

On vérifie pour tout $1 \leq i \leq n$,

$$\frac{f(\mathbf{x}_i) - \hat{f}_{-i(\mathbf{x}_i)}}{s_{-i}(\mathbf{x}_i)} \in [-3, 3]\,.$$

Plan

Paradigme bayésien

On souhaite effectuer une estimation ensembliste de

$$t_\rho = \mathbb{P}_{\mathbf{X}}(f(\mathbf{X}) < \rho) = \int_{\mathbb{R}^d} \mathbb{I}_{\{f(\mathbf{x})<\rho\}} g(\mathbf{x}) \mathrm{d}\mathbf{x} = \mathbb{P}_\omega \left(\{\omega | f(\mathbf{X}(\omega)) < \rho\}\right) .$$

- Le processus $(Y_{\mathbf{x}})_{\mathbf{x}\in E}$ est une loi a priori pour $f(\cdot)$.
- On suppose que le processus $(Y_{\mathbf{x}})_{\mathbf{x}\in E}$ est indépendant de $\mathbf{X}$.
- On considère ainsi que $((Y_{\mathbf{x}})_{\mathbf{x}\in E}, \mathbf{X})$ est une application mesurable de $(\Omega \times \Gamma, \mathcal{A} \times \mathcal{B}, \mathbb{P}_\omega \times \mathbb{P}_\gamma)$ sur $(F \times E, \mathcal{F} \times \mathcal{B}_E, \mathbb{P}_{Y_{\mathbf{x}}} \times \mathbb{P}_{\mathbf{X}})$.

Dans ce cadre bayésien, t_ρ est une variable aléatoire :

$$t_\rho = \mathbb{P}_{\mathbf{X}} \left(Y_{\mathbf{X}} < \rho\right) .$$

Objectif : construire un intervalle de crédibilité sur t_ρ. On cherche la valeur $0 < a < 1$ fixée telle que la probabilité que $t_\rho < a$ soit égale à $1 - \alpha$ (avec α petit).

Estimateur

L'estimation bayésienne de t_ρ associée à la fonction de perte quadratique est donnée par la moyenne de la loi a posteriori de t_ρ :

$$\begin{aligned}
\mathbb{E}_{(Y_{\mathbf{x}})_{\mathbf{x}\in E}}\left(t_\rho | y_{\mathbf{x}_1}, \ldots, y_{\mathbf{x}_n}\right) &= \mathbb{E}_{(Y_{\mathbf{x}})_{\mathbf{x}\in E}}\left(\mathbb{P}_{\mathbf{X}}\left(Y_{\mathbf{X}} < \rho\right) | y_{\mathbf{x}_1}, \ldots, y_{\mathbf{x}_n}\right) \\
&= \mathbb{E}_{(Y_{\mathbf{x}})_{\mathbf{x}\in E}}\left(\mathbb{E}_{\mathbf{X}}\left(\mathbb{I}_{Y_{\mathbf{X}}<\rho}\right) | y_{\mathbf{x}_1}, \ldots, y_{\mathbf{x}_n}\right) \\
&= \mathbb{E}_{\mathbf{X}}\left(\mathbb{E}_{(Y_{\mathbf{x}})_{\mathbf{x}\in E}}\left(\mathbb{I}_{Y_{\mathbf{X}}<\rho} | y_{\mathbf{x}_1}, \ldots, y_{\mathbf{x}_n}\right)\right) \\
&= \mathbb{E}_{\mathbf{X}}\left(\mathbb{E}_{(Y^D_{\mathbf{x}})_{\mathbf{x}\in E}}\left(\mathbb{I}_{Y^D_{\mathbf{X}}<\rho}\right)\right) \\
&= \mathbb{E}_{\mathbf{X}}\left(\mathbb{P}\left(Y^D_{\mathbf{X}} < \rho\right)\right) \\
&= \mathbb{E}_{\mathbf{X}}\left(\Phi\left(\frac{\rho - \hat{f}(\mathbf{X}; y_{\mathbf{x}_1}, \ldots, y_{\mathbf{x}_n})}{\sqrt{\mathrm{MSE}(\mathbf{X})}}\right)\right),
\end{aligned}$$

où Φ est le fonction de répartition de la loi normale centrée réduite.

Calcul de l'estimateur

Une intégration numérique sur la variable **X** est possible par une méthode de Monte-Carlo. Il n'est pas nécessaire d'appeler la fonction f.

Remarque

D'après l'inégalité Markov, nous avons

$$\mathbb{P}\left(t_\rho < a | y_{\mathbf{x}_1}, \ldots, y_{\mathbf{x}_n}\right) \geq 1 - \mathbb{E}_{(Y_{\mathbf{x}})_{\mathbf{x}\in E}}\left(t_\rho | y_{\mathbf{x}_1}, \ldots, y_{\mathbf{x}_n}\right) / a \,.$$

Ainsi avec une probabilité supérieure à $1 - \alpha$, nous avons

$$t_\rho < \mathbb{E}_{(Y_{\mathbf{x}})_{\mathbf{x}\in E}}\left(t_\rho | y_{\mathbf{x}_1}, \ldots, y_{\mathbf{x}_n}\right) / \alpha \,.$$

Réalisations de t_ρ

Nous pouvons simuler des réalisations de t_ρ suivant sa loi a posteriori.

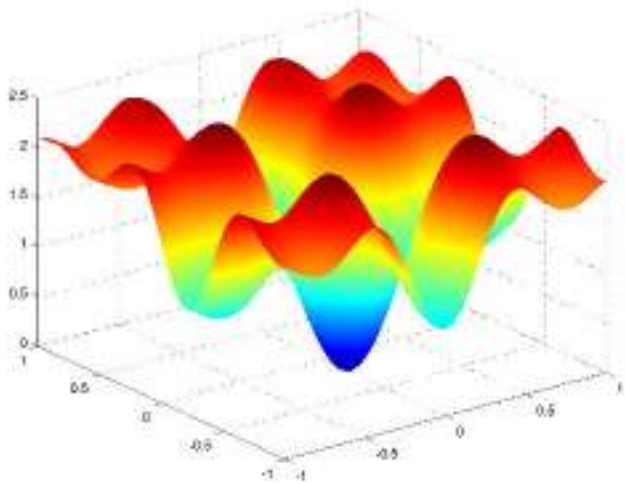
Nous utilisons ces trois étapes pour obtenir une réalisation de $t_\rho = \mathbb{P}_{\mathbf{X}}(Y_{\mathbf{X}} < \rho)$:

1. **Simulation de points du champ gaussien $Y_\cdot^D$:** $(y_{\tilde{x}_i})_{1 \leq i \leq \tilde{n}}$ est simulé suivant la loi du processus a posteriori $Y_\cdot^D$ où les points $\tilde{x}_1, \ldots, \tilde{x}_n$ forment une grille dans E.
2. **Reconstruction du champ gaussien :** Grâce à une méthode de krigeage, nous interpolons les points $(y_{\tilde{x}_i})_{1 \leq i \leq \tilde{n}} \cup (y_{x_i})_{1 \leq i \leq n}$ c'est à dire les réalisations du processus a posteriori et les évalutions de la fonction f. Nous considérons que cette interpolation est la réalisation du processus $Y_\cdot^D$ sur E.
3. **Intégration numérique :** Nous pouvons alors calculer t_ρ pour cette trajectoire grâce à une méthode de Monte-Carlo massive sur $\mathbf{X}$.

Essai sur une fonction jouet

On utilise la fonction $f : [-10, 10]^2 \to \mathbb{R}_+$ suivante :

$$f(x_1, x_2) = -\frac{\sin(x_1)}{x_1} - \frac{\sin(x_2 + 2)}{x_2 + 2} + 2;$$

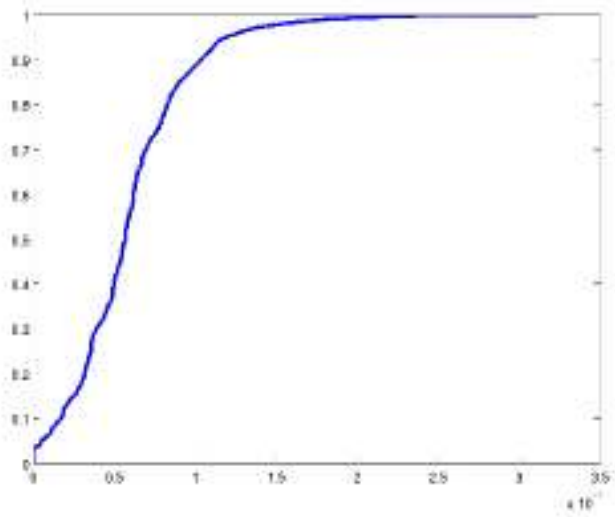


On veut estimer la probabilité $t_\rho = \mathbb{P}(f(\mathbf{X}) < \rho)$ pour $\rho = 0.01$.
Par Monte-Carlo intensif, on obtient $t_\rho = \mathbb{P}(f(\mathbf{X}) < \rho) = 4.7 \cdot 10^{-4}$.

Exemple $n = 100$

1. Moyenne estimée : $5.3 \cdot 10^{-4}$ avec 10^5 copies de $\mathbf{X}$.

2. $\tilde{n} = 50$, 10^4 trajectoires simulées, 10^5 copies de $\mathbf{X}$ simulées.



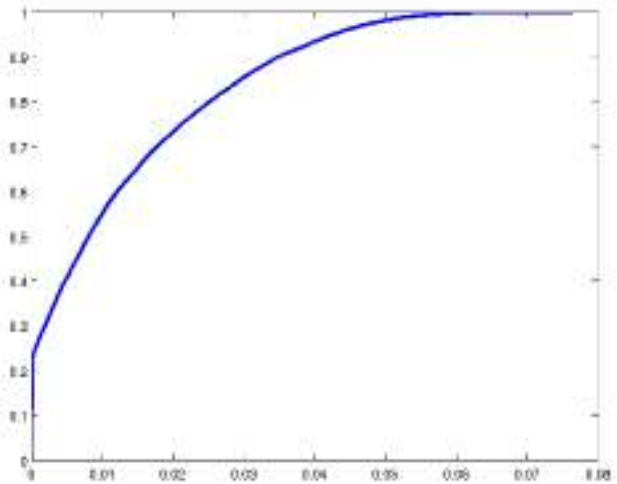
- Moyenne estimée : $5.9 \cdot 10^{-4}$,
- Variance estimée : $1.3 \cdot 10^{-7}$,
- Quantile à 99% : $1.8 \cdot 10^{-3}$.

L'estimateur de Monte-Carlo proposait au mieux l'intervalle de confiance $[0; 0.045]$.

Exemple $n = 20$

1. Moyenne estimée : $1.3 \cdot 10^{-2}$ avec 10^5 copies de $\mathbf{X}$.

2. $\tilde{n} = 144$, 10^4 trajectoires simulées, 10^5 copies de $\mathbf{X}$ simulées.



- Moyenne estimée : $1.3 \cdot 10^{-2}$,
- Variance estimée : $2.0 \cdot 10^{-4}$,
- Quantile à 99% : $5.4 \cdot 10^{-2}$.

L'estimateur de Monte-Carlo proposait au mieux l'intervalle de confiance $[0; 0.21]$.

Plan

Echantillonnage préférentiel grâce au modèle réduit

- Modèle réduit construit sur $N/2$ points : $\hat{f}$.
- On note $\widehat{R} = \{\mathbf{x} \ : \ \hat{f}(\mathbf{x}) < \rho - 3\sqrt{\mathrm{MSE}(\mathbf{x})}\}$.
- $\mathbb{P}_{\mathbf{X}}(\mathbf{X} \in \widehat{R})$ est calculée grâce à un Monte-Carlo intensif.
- On propose d'utiliser la distribution instrumentale

$$h : \mathbf{x} \mapsto \mathbb{1}_{\widehat{R}}(\mathbf{x}) g(\mathbf{x}) \mathbb{P}(\mathbf{X} \in \widehat{R})^{-1} .$$

Si $(Z_1, \ldots, Z_{N/2})$ est un $N/2$-échantillon simulé suivant h,

$$\hat{t}_\rho^{IS} = \frac{2}{N} \sum_{i=1}^{N/2} \mathbb{I}_R(Z_i) \frac{g(Z_i)}{h(Z_i)} = \frac{2\mathbb{P}(\mathbf{X} \in \widehat{R})}{N} \sum_{i=1}^{N/2} \mathbb{I}_R(Z_i) .$$

Propriétés de l'estimateur

- $\mathbb{E}_{\mathbf{X}}(\hat{t}_\rho^{IS}) = \mathbb{P}(\mathbf{X} \in \widehat{R} \cap R)$
- $\mathrm{Var}_{\mathbf{X}}(\hat{t}_\rho^{IS}) = \frac{2}{N}\mathbb{P}(\mathbf{X} \in \widehat{R} \cap R)(\mathbb{P}(\mathbf{X} \in \widehat{R}) - \mathbb{P}(\mathbf{X} \in \widehat{R} \cap R))$

Notre estimateur souffre d'un biais négatif

$$\text{biais} = \mathbb{P}_{\mathbf{X}}(\mathbf{X} \in R \cap \widehat{R}) - \mathbb{P}_{\mathbf{X}}(\mathbf{X} \in R) = -\mathbb{P}_{\mathbf{X}}(\mathbf{X} \in R \cap \mathbf{X} \in \widehat{R}^C)\,.$$

Inégalité de concentration

Avec probabilité $1 - \delta$

$$t_\rho \leq b_{\delta,N}(\hat{t}_\rho^{IS}) - \text{biais}. \tag{1}$$

Il faut contrôler le biais !

Contrôle du biais

f est considérée comme une réalisation du processus gaussien $Y_\cdot$ pour un ω_0 donné, dont on a obversé les valeurs $y_1, \ldots, y_{N/2}$ prises en des points $\mathbf{x}_1 \ldots, \mathbf{x}_{N/2}$. Ainsi

$$\forall \mathbf{x}, \quad f(\mathbf{x}) = Y_{\mathbf{x}}(\omega_0) = Y_{\mathbf{x}}^D(\omega_0)\,.$$

On définit la variable aléatoire $B : (\omega, \mathcal{A}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$

$$B = \mathbb{P}_{\mathbf{X}}\left(\mathbf{X} \in R \cap Y_{\mathbf{X}}^D < \hat{f}(\mathbf{X}) - 3\sqrt{\mathrm{MSE}(\mathbf{X})}\right) = \mathbb{E}_{\mathbf{X}}\left(\mathbb{I}_{\mathbf{X}\in R}\mathbb{I}_{Y_{\mathbf{X}}^D < \hat{f}(\mathbf{X}) - 3\sqrt{\mathrm{MSE}(\mathbf{X})}}\right)\,.$$

$$\mathbb{E}_{(Y_{\mathbf{x}}^D)_{\mathbf{x}\in E}}(B) = 0.23\% \cdot \mathbb{P}_{\mathbf{X}}(\mathbf{X} \in R)\,.$$

Par l'inégalité de Markov, B étant presque sûrement positive ou nulle, on a

$$\mathbb{P}_{(Y_{\mathbf{x}}^D)_{\mathbf{x}\in E}}(B \geq 1/2\mathbb{P}_{\mathbf{X}}(\mathbf{X} \in R)) \leq \frac{0.23\%\mathbb{P}_{\mathbf{X}}(\mathbf{X} \in R)}{1/2\mathbb{P}_{\mathbf{X}}(\mathbf{X} \in R)} = 0.46\%\,.$$

On rappelle que R est défini ainsi $R = \{\mathbf{x} : f(\mathbf{x}) < \rho\} = \{\mathbf{x} : Y_{\mathbf{x}}^D(\omega_0) < \rho\}$. On a alors l'inégalité,

$$\begin{aligned}
\mathbb{P}_{\mathbf{X}}\left(\mathbf{X} \in R \cap \rho < \hat{f}(\mathbf{X}) - 3\sqrt{\text{MSE}(\mathbf{X})}\right) &\leq \mathbb{P}_{\mathbf{X}}\left(\mathbf{X} \in R \cap f(\mathbf{X}) < \hat{f}(\mathbf{X}) - 3\sqrt{\text{MSE}(\mathbf{X})}\right) \\
\Leftrightarrow \mathbb{P}_{\mathbf{X}}\left(\mathbf{X} \in R \cap \mathbf{X} \in \widehat{R}^C\right) &\leq \mathbb{P}_{\mathbf{X}}\left(\mathbf{X} \in R \cap Y_{\mathbf{X}}(\omega_0) < \hat{f}(\mathbf{X}) - 3\sqrt{\text{MSE}(\mathbf{X})}\right) \\
\Leftrightarrow \mathbb{P}_{\mathbf{X}}\left(\mathbf{X} \in R \cap \mathbf{X} \in \widehat{R}^C\right) &\leq B(\omega_0).
\end{aligned}$$

Avec probabilité 99.54%, $B \leq 1/2\mathbb{P}_{\mathbf{X}}(\mathbf{X} \in R)$, donc on a

Avec une confiance de 99.54%,

$$-\text{biais} = \mathbb{P}_{\mathbf{X}}\left(\mathbf{X} \in R \cap \mathbf{X} \in \widehat{R}^C\right) \leq 1/2\mathbb{P}_{\mathbf{X}}(\mathbf{X} \in R)\,.$$

Conclusion

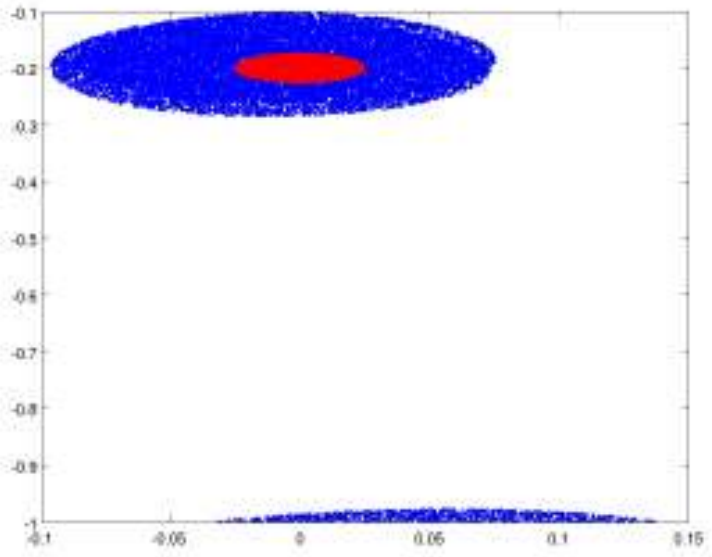
Par conséquent, il y a deux niveaux de confiance

1. Avec probabilité 99.54%, $-\text{biais} \leq 1/2\mathbb{P}_{\mathbf{X}}(\mathbf{X} \in R) = 1/2t_\rho$,
2. le cas échéant, avec probabilité $1-\delta$, d'après l'inégalité (1),

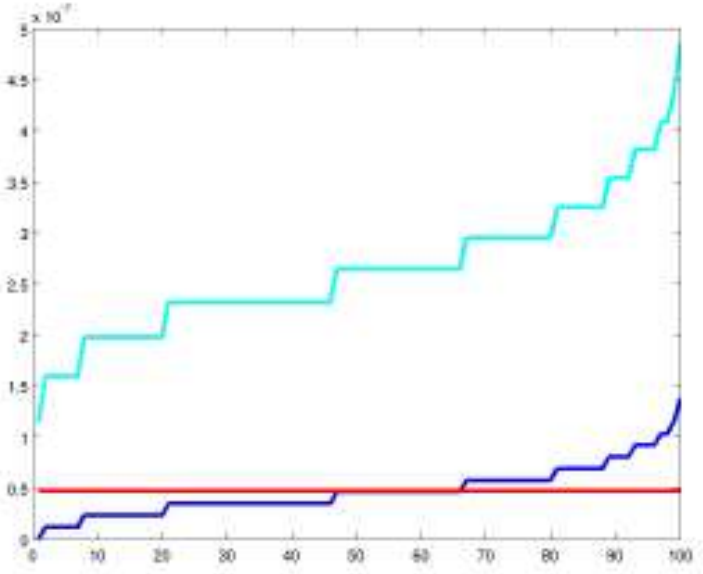
$$\begin{aligned} t_\rho \leq b_{\delta,N}(\hat{t}_\rho^{IS}) - \text{biais} &\leq b_{\delta,N}(\hat{t}_\rho^{IS}) + 1/2t_\rho \\ t_\rho &\leq 2b_{\delta,N}(\hat{t}_\rho^{IS}). \end{aligned}$$

100 estimations

50 points pour construire le modèle réduit, 50 points pour l'estimateurs $\hat{t}_\rho^{IS}$.



En rouge le domaine R,
En bleu le domaine $\widehat{R}$.



En rouge la probabilité $\mathbb{P}(\mathbf{X} \in R)$,
En bleu les estimations $\hat{t}_\rho^{IS}$
En cyan, les bornes supérieures.

Conclusion

On a présenté 2 méthodes pour estimer la probabilité d'un événement rare :

1. **Estimation bayésienne :**
   - Une seule série d'appels à f ;
   - Richesse donnée par la loi a posteriori ;
   - Résulats intéressants avec peu de points tant que le modèle a priori est "validé" ;
   - Simulation coûteuse de processus en "grande" dimension ;
2. **Échantillonnage préférentiel :**
   - Simulation de l'échantillon simple (possible en dimension plus grande) ;
   - Deux niveaux de confiance ;
   - Deux séries d'appels à f ;