

Propagation d'incertitudes et estimation d'un quantile

Tatiana Labopin-Richard

04 septembre 2013

Introduction

- Stage de M2R effectué à l'IMT et encadré par Fabrice Gamboa et Aurélien Garivier.

Introduction

- Stage de M2R effectué à l'IMT et encadré par Fabrice Gamboa et Aurélien Garivier.
- L'entreprise Orange souhaite connaître des informations sur le taux d'ondes électromagnétiques qu'absorbe un fœtus (quantile).

Introduction

- Stage de M2R effectué à l'IMT et encadré par Fabrice Gamboa et Aurélien Garivier.
- L'entreprise Orange souhaite connaître des informations sur le taux d'ondes électromagnétiques qu'absorbe un fœtus (quantile).
- Ce taux dépend de plusieurs paramètres et chaque évaluation de ce taux peut être faite quasiment sans erreur mais coûte très cher : on veut une bonne approximation du quantile en faisant très peu de mesures (budget de 100).

Plan

- 1 Modélisation du problème
 - Modèle
 - Rappels sur le krigeage
- 2 Choix des estimateurs
- 3 Stratégie de choix du design : la méthode SUR
 - Vers une méthode d'apprentissage actif
 - Sélection du nouveau point à l'étape n : méthode SUR
 - Algorithme
- 4 Une autre méthode venant de l'optimisation bayésienne
 - Idée
 - A la recherche d'une forme analytique de notre critère
 - Algorithme
- 5 Résultats
 - En dimension 2

- 1 Modélisation du problème
 - Modèle
 - Rappels sur le krigeage
- 2 Choix des estimateurs
- 3 Stratégie de choix du design : la méthode SUR
 - Vers une méthode d'apprentissage actif
 - Sélection du nouveau point à l'étape n : méthode SUR
 - Algorithme
- 4 Une autre méthode venant de l'optimisation bayésienne
 - Idée
 - A la recherche d'une forme analytique de notre critère
 - Algorithme
- 5 Résultats
 - En dimension 2
 - Dimension 5

Modèle

On adopte le modèle suivant :

$$Y = f(X)$$

avec $X \sim \mathcal{U}[0, 1]^d$ le vecteur des paramètres et f une fonction inconnue mais que l'on sait évaluer.

Modèle

On adopte le modèle suivant :

$$Y = f(X)$$

avec $X \sim \mathcal{U}[0, 1]^d$ le vecteur des paramètres et f une fonction inconnue mais que l'on sait évaluer.

BUT : Trouver une stratégie pour obtenir la meilleure approximation d'un quantile de Y en n'évaluant f seulement qu'en 100 points.

Modèle

On adopte le modèle suivant :

$$Y = f(X)$$

avec $X \sim \mathcal{U}[0, 1]^d$ le vecteur des paramètres et f une fonction inconnue mais que l'on sait évaluer.

BUT : Trouver une stratégie pour obtenir la meilleure approximation d'un quantile de Y en n'évaluant f seulement qu'en 100 points.

2 problèmes :

- Choix d'un bon estimateur

Modèle

On adopte le modèle suivant :

$$Y = f(X)$$

avec $X \sim \mathcal{U}[0, 1]^d$ le vecteur des paramètres et f une fonction inconnue mais que l'on sait évaluer.

BUT : Trouver une stratégie pour obtenir la meilleure approximation d'un quantile de Y en n'évaluant f seulement qu'en 100 points.

2 problèmes :

- Choix d'un bon estimateur
- Choix d'une bonne stratégie de choix de points à évaluer

- 1 Modélisation du problème
 - Modèle
 - Rappels sur le krigeage
- 2 Choix des estimateurs
- 3 Stratégie de choix du design : la méthode SUR
 - Vers une méthode d'apprentissage actif
 - Sélection du nouveau point à l'étape n : méthode SUR
 - Algorithme
- 4 Une autre méthode venant de l'optimisation bayésienne
 - Idée
 - A la recherche d'une forme analytique de notre critère
 - Algorithme
- 5 Résultats
 - En dimension 2
 - Dimension 5

Krigeage

Hypothèse classique : f est une réalisation d'un processus gaussien (Y) centré de fonction de covariance fixée k .

Avantage : La loi a posteriori reste gaussienne et est connue.

Propriété

Soit $Y \sim PG(0, k)$. Soit $Y_n = (Y_1, \dots, Y_n)$ vecteurs d'évaluation en $X_n = (x_1, \dots, x_n)$. Alors la loi de $Y(x)$ sachant la tribu \mathcal{F}_n pour $x \in \mathcal{X}$ est gaussienne de paramètres :

$$\begin{aligned} m_n(x) &= k_n(x)^T K_n^{-1} y_n \\ c_n(x, x') &= k(x, x') - k_n(x)^T K_n^{-1} k_n(x') \end{aligned}$$

avec $k_n(x) = [k(x_1, x), \dots, k(x_n, x)]'$ et $K_n = [k(x_i, x_j)]_{1 \leq i, j \leq n}$, et dans la suite nous noterons $c_n(x, x) = s_n^2(x)$.

Formules de mise à jour

Inconvénient : Pour tout calcul de loi a posteriori, il faut faire une inversion de matrice.

Formules de mise à jour

Inconvénient : Pour tout calcul de loi a posteriori, il faut faire une inversion de matrice.

Apprentissage actif : Nous allons chercher une stratégie qui choisira les nouveaux points à évaluer un par un.

Formules de mise à jour

Inconvénient : Pour tout calcul de loi a posteriori, il faut faire une inversion de matrice.

Apprentissage actif : Nous allons chercher une stratégie qui choisira les nouveaux points à évaluer un par un.

Formules de mise à jour à 1 pas

$$\begin{aligned}
 m_{n+1}(x) &= m_n(x) + \frac{c_n(x_{n+1}, x)}{s_n^2(x_{n+1})} (Y(x_{n+1}) - m_n(x_{n+1})) \\
 s_{n+1}^2(x) &= s_n^2(x) - \frac{c_n^2(x_{n+1}, x)}{s_n^2(x_{n+1})} \\
 c_{n+1}(x, y) &= c_n(x, y) - \frac{c_n(x_{n+1}, x)c_n(x_{n+1}, y)}{s_n^2(x_{n+1})}
 \end{aligned}$$

- 1 Modélisation du problème
 - Modèle
 - Rappels sur le krigeage
- 2 Choix des estimateurs
- 3 Stratégie de choix du design : la méthode SUR
 - Vers une méthode d'apprentissage actif
 - Sélection du nouveau point à l'étape n : méthode SUR
 - Algorithme
- 4 Une autre méthode venant de l'optimisation bayésienne
 - Idée
 - A la recherche d'une forme analytique de notre critère
 - Algorithme
- 5 Résultats
 - En dimension 2
 - Dimension 5

Estimateurs de la probabilité d'échec

Notations : On notera \mathbb{P}_n et \mathbb{E}_n les probabilités et espérance conditionnellement à la tribu \mathcal{F}_n

Estimateurs de la probabilité d'échec

Notations : On notera \mathbb{P}_n et \mathbb{E}_n les probabilités et espérance conditionnellement à la tribu \mathcal{F}_n

Probabilité d'échec : dans la littérature (Bect, Chevalier, Ginsbourger, Li, Picheny, Vazquez), $\mathbb{P}(f(X) > u) := \alpha(f)$ souvent étudiée par ces méthodes.

Estimateurs proposés :

- $\hat{\alpha}_n^1 = \alpha(m_n)$
- $\hat{\alpha}_n^2 = \mathbb{E}_n[\alpha(Y)] = \mathbb{E}_n[\int_{\mathbb{X}} 1_{Y>u} dP_{\mathbb{X}}]$

Etude difficile de ces estimateurs.

Adaptation de ces estimateurs à notre problème

Deux estimateurs pour notre problème :

On appelle $q(f)$ le quantile d'ordre α de la distribution $f(X)$.

- $\hat{q}_n^1 = q(m_n)$
- $\hat{q}_n^2 = \mathbb{E}_n[q(Y)]$

Et pour trouver $q(m_n)$ où $q(Y)$ nous utilisons le quantile empirique (noté \hat{q}_n^*).

- $\hat{q}_n^1 = \hat{q}_n^*(m_n)$
- $\hat{q}_n^2 = \mathbb{E}_n[\hat{q}_n^*(Y)]$

- 1 Modélisation du problème
 - Modèle
 - Rappels sur le krigeage
- 2 Choix des estimateurs
- 3 **Stratégie de choix du design : la méthode SUR**
 - **Vers une méthode d'apprentissage actif**
 - Sélection du nouveau point à l'étape n : méthode SUR
 - Algorithme
- 4 Une autre méthode venant de l'optimisation bayésienne
 - Idée
 - A la recherche d'une forme analytique de notre critère
 - Algorithme
- 5 Résultats
 - En dimension 2
 - Dimension 5

Apprentissage actif

Première idée : Random \mapsto trop de temps dans des endroits inutiles.

Apprentissage actif

Première idée : Random \mapsto trop de temps dans des endroits inutiles.

Stratégie d'apprentissage actif : on choisit les points un par un en utilisant toute l'information sur f disponible grâce aux précédentes évaluations (krigeage).

Apprentissage actif

Première idée : Random \mapsto trop de temps dans des endroits inutiles.

Stratégie d'apprentissage actif : on choisit les points un par un en utilisant toute l'information sur f disponible grâce aux précédentes évaluations (krigeage).

Deuxième idée : Méthode GPExplor : on choisit le point qui minimise l'incertitude sur f :

$$x_{n+1} = \operatorname{argmax}_{x \in \mathbb{X}} s_n^2(x)$$

\mapsto Il existe mieux : compromis exploration, exploitation.

Plan de la stratégie

- 1) On tire une grille de l'espace $[0, 1]^d$ (\mathcal{X}).
- 2) Choix d'une sous-grille \mathcal{X}_0 . Evaluation des points de \mathcal{X}_0 et calcul de l'estimateur courant ($\hat{q}(f)^0$) avec le modèle substitut. \mathcal{X}_0 , leur évaluation et $\hat{q}(f)^0$ engendrent la tribu \mathcal{F}_0 .
- 3) A chaque étape n (allant de n_0 à n_{budget}) on cherche un nouveau point x_n à évaluer, en utilisant seulement la tribu \mathcal{F}_{n-1} , on évalue ce nouveau point et on met à jour notre estimateur (et donc la tribu).

Plan de la stratégie

- 1) On tire une grille de l'espace $[0, 1]^d$ (\mathcal{X}).
- 2) Choix d'une sous-grille \mathcal{X}_0 . Evaluation des points de \mathcal{X}_0 et calcul de l'estimateur courant ($\hat{q}(f)^0$) avec le modèle substitut. \mathcal{X}_0 , leur évaluation et $\hat{q}(f)^0$ engendrent la tribu \mathcal{F}_0 .
- 3) A chaque étape n (allant de n_0 à n_{budget}) on cherche un nouveau point x_n à évaluer, en utilisant seulement la tribu \mathcal{F}_{n-1} , on évalue ce nouveau point et on met à jour notre estimateur (et donc la tribu).

Point difficile : Comment trouver le nouveau point à évaluer ?

- 1 Modélisation du problème
 - Modèle
 - Rappels sur le krigeage
- 2 Choix des estimateurs
- 3 **Stratégie de choix du design : la méthode SUR**
 - Vers une méthode d'apprentissage actif
 - **Sélection du nouveau point à l'étape n : méthode SUR**
 - Algorithme
- 4 Une autre méthode venant de l'optimisation bayésienne
 - Idée
 - A la recherche d'une forme analytique de notre critère
 - Algorithme
- 5 Résultats
 - En dimension 2
 - Dimension 5

Article de 2012 de Marjorie Jala, Céline Lévy-Leduc, Eric Moulines, Emmanuelle Conil et Joe Wiart.

Article de 2012 de Marjorie Jala, Céline Lévy-Leduc, Eric Moulines, Emmanuelle Conil et Joe Wiart.

Méthode SUR : Stepwise Uncertainly Réduction : à chaque pas, on veut réduire "l'incertitude".

$$x_{n+1}^* = \min_{x_{n+1} \in \mathcal{X}} V_n(x_{n+1})$$

$$V_n(x_{n+1}) = \int \text{Var}(\hat{q}(f) | \mathcal{F}_n^{(x_{n+1}, y)}) \phi_{(m_n(x_{n+1}), s_n^2(x_{n+1}))}(y) dy$$

où $\phi_{(m_n(x_{n+1}), s_n^2(x_{n+1}))}$ est la densité gaussienne.

Article de 2012 de Marjorie Jala, Céline Lévy-Leduc, Eric Moulines, Emmanuelle Conil et Joe Wiart.

Méthode SUR : Stepwise Uncertainly Réduction : à chaque pas, on veut réduire "l'incertitude".

$$x_{n+1}^* = \min_{x_{n+1} \in \mathcal{X}} V_n(x_{n+1})$$

$$V_n(x_{n+1}) = \int \text{Var}(\hat{q}(f) | \mathcal{F}_n^{(x_{n+1}, y)}) \phi_{(m_n(x_{n+1}), s_n^2(x_{n+1}))}(y) dy$$

où $\phi_{(m_n(x_{n+1}), s_n^2(x_{n+1}))}$ est la densité gaussienne.

Problème : Pas de forme analytique pour $V \mapsto$ grosse complexité de calcul (3 boucles Monte Carlo imbriquées).

- 1 Modélisation du problème
 - Modèle
 - Rappels sur le krigeage
- 2 Choix des estimateurs
- 3 Stratégie de choix du design : la méthode SUR**
 - Vers une méthode d'apprentissage actif
 - Sélection du nouveau point à l'étape n : méthode SUR
 - Algorithme**
- 4 Une autre méthode venant de l'optimisation bayésienne
 - Idée
 - A la recherche d'une forme analytique de notre critère
 - Algorithme
- 5 Résultats
 - En dimension 2
 - Dimension 5

Algorithme

1. Evaluer la distribution postérieure Y_{n_0} par les formules de krigeage ($Y_{n_0} \sim PG(\mu_{n_0}, k_{n_0})$).
2. Estimer le quantile dans ce modèle substitut.
3. Simuler un N -échantillon de Y_{n_0} : $(Y_{n_0}^{(1)}, Y_{n_0}^{(2)}, \dots, Y_{n_0}^{(N)})$.
4. Estimation du critère pour tout $x_{n+1} \in \mathcal{X}/\mathcal{X}_0$: boucle sur les x_{n+1} .
5. Minimiser sur les x_{n+1} le critère $V_n(x_{n+1})$.
6. Evaluer le point qui minimisait et grossir la tribu des observations.
7. Retourner à la première étape si le budget d'évaluations n'est pas atteint.

Détails de l'étape 4

A. Pour i allant de 1 à N :

- Evaluer la distribution postérieure de $Y_{n_0}^{(x_{n+1}, Y_{n_0}^i(x_{n+1}))}$.
- Simuler un N' -échantillon $(Y_{n_0}^{(1)}(x_{n+1}, Y_{n_0}^i(x_{n+1})), \dots, Y_{n_0}^{(N')}(x_{n+1}, Y_{n_0}^i(x_{n+1})))$.
- Pour j allant de 1 à N' :
 - Estimer le quantile $\hat{q}_{n_0}^{(j)}(Y_{n_0}^{(j)}(x, \varepsilon_{n_0}^i(x)))$.
- Calculer la variance empirique des $\hat{q}_{n_0}^{(i)}(x_{n+1})$.

B. Evaluer $V_n(x_{n+1}) = \frac{1}{N} \sum_{i=1}^N s_{n(i)}(x_{n+1})$.

- 1 Modélisation du problème
 - Modèle
 - Rappels sur le krigeage
- 2 Choix des estimateurs
- 3 Stratégie de choix du design : la méthode SUR
 - Vers une méthode d'apprentissage actif
 - Sélection du nouveau point à l'étape n : méthode SUR
 - Algorithme
- 4 Une autre méthode venant de l'optimisation bayésienne
 - Idée
 - A la recherche d'une forme analytique de notre critère
 - Algorithme
- 5 Résultats
 - En dimension 2
 - Dimension 5

Idée d'optimisation

Idée de Bect, Chevalier, Ginsbouger, Picheny, Vasquez pour estimer le minimum de f :

Idée d'optimisation

Idée de Bect, Chevalier, Ginsbouger, Picheny, Vasquez pour estimer le minimum de f :

$$\Gamma_n = \int_{\mathbb{X}} \mathbb{P}_n(Y(x) \leq y_{min}^n) dx$$

où y_{min}^n représente l'estimateur du minimum au temps n .

Idée d'optimisation

Idée de Bect, Chevalier, Ginsbouger, Picheny, Vasquez pour estimer le minimum de f :

$$\Gamma_n = \int_{\mathbb{X}} \mathbb{P}_n(Y(x) \leq y_{min}^n) dx$$

où y_{min}^n représente l'estimateur du minimum au temps n .

Stratégie : on recherche le point tel que si on le choisit à l'étape $(n+1)$, on aura en moyenne la plus petite probabilité de se tromper :

$$x_{n+1}^* = \underset{\mathbb{X}}{\operatorname{argmin}} \Gamma_{n+1}(x_{n+1}) = \underset{\mathbb{X}}{\operatorname{argmin}} \int_{\mathbb{X}} \mathbb{P}_n(Y(x) \leq \min(y_{min}^n, y_{n+1}) | Y_{n+1} = Y(x_{n+1})) dx$$

Idée d'optimisation

Idée de Bect, Chevalier, Ginsbougner, Picheny, Vasquez pour estimer le minimum de f :

$$\Gamma_n = \int_{\mathbb{X}} \mathbb{P}_n(Y(x) \leq y_{min}^n) dx$$

où y_{min}^n représente l'estimateur du minimum au temps n .

Stratégie : on recherche le point tel que si on le choisit à l'étape $(n+1)$, on aura en moyenne la plus petite probabilité de se tromper :

$$x_{n+1}^* = \underset{\mathbb{X}}{\operatorname{argmin}} \Gamma_{n+1}(x_{n+1}) = \underset{\mathbb{X}}{\operatorname{argmin}} \int_{\mathbb{X}} \mathbb{P}_n(Y(x) \leq \min(y_{min}^n, y_{n+1}) | Y_{n+1} = Y(x_{n+1})) dx$$

Intérêt : La probabilité dans l'intégrale a une forme analytique en fonction de la fonction de répartition d'un vecteur gaussien

Comment utiliser cette idée sur notre problème

Le quantile est défini ainsi : le réel q_α tel que $\mathbb{P}(Y \geq q_\alpha) = 1 - \alpha$.
Nous retrouvons donc les probabilités qui ont une forme analytique.

Comment utiliser cette idée sur notre problème

Le quantile est défini ainsi : le réel q_α tel que $\mathbb{P}(Y \geq q_\alpha) = 1 - \alpha$.
Nous retrouvons donc les probabilités qui ont une forme analytique.

Nous créons alors un autre critère à minimiser, qui mesure en moyenne l'erreur commise par notre estimateur :

$$\Gamma_n = \int_{\mathbb{X}} \left| \mathbb{P}_n(Y(x) \geq \hat{q}_\alpha^n) - (1 - \alpha) \right| dx$$

Comment utiliser cette idée sur notre problème

Le quantile est défini ainsi : le réel q_α tel que $\mathbb{P}(Y \geq q_\alpha) = 1 - \alpha$.
Nous retrouvons donc les probabilités qui ont une forme analytique.

Nous créons alors un autre critère à minimiser, qui mesure en moyenne l'erreur commise par notre estimateur :

$$\Gamma_n = \int_{\mathbb{X}} \left| \mathbb{P}_n(Y(x) \geq \hat{q}_\alpha^n) - (1 - \alpha) \right| dx$$

Ainsi, au début de l'étape $(n + 1)$ il nous faudra choisir le x_{n+1} qui minimisera le critère de l'étape $(n + 1)$ si l'on avait choisit ce point d'évaluation : $x_{n+1}^* = \underset{\mathbb{X}}{\operatorname{argmin}} \Gamma_{n+1}(x_{n+1})$.

Comment utiliser cette idée sur notre problème

Le quantile est défini ainsi : le réel q_α tel que $\mathbb{P}(Y \geq q_\alpha) = 1 - \alpha$.
Nous retrouvons donc les probabilités qui ont une forme analytique.

Nous créons alors un autre critère à minimiser, qui mesure en moyenne l'erreur commise par notre estimateur :

$$\Gamma_n = \int_{\mathbb{X}} \left| \mathbb{P}_n(Y(x) \geq \hat{q}_\alpha^n) - (1 - \alpha) \right| dx$$

Ainsi, au début de l'étape $(n + 1)$ il nous faudra choisir le x_{n+1} qui minimisera le critère de l'étape $(n + 1)$ si l'on avait choisit ce point d'évaluation : $x_{n+1}^* = \underset{\mathbb{X}}{\operatorname{argmin}} \Gamma_{n+1}(x_{n+1})$.

Nous cherchons alors une forme quasi-analytique de notre critère :

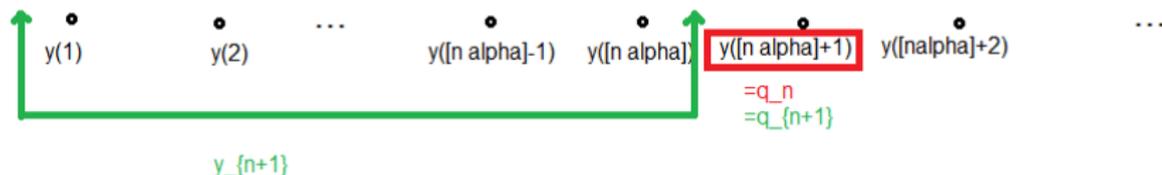
$$\Gamma_{n+1}(x_{n+1}) = \int_{\mathbb{X}} \left| \mathbb{P}_n(Y(x) \geq \hat{q}_\alpha^{n+1} | Y(x_{n+1}) = Y_{n+1}) - (1 - \alpha) \right| dx$$

- 1 Modélisation du problème
 - Modèle
 - Rappels sur le krigeage
- 2 Choix des estimateurs
- 3 Stratégie de choix du design : la méthode SUR
 - Vers une méthode d'apprentissage actif
 - Sélection du nouveau point à l'étape n : méthode SUR
 - Algorithme
- 4 Une autre méthode venant de l'optimisation bayésienne
 - Idée
 - A la recherche d'une forme analytique de notre critère
 - Algorithme
- 5 Résultats
 - En dimension 2
 - Dimension 5

Mise à jour du quantile

Nous voulons exprimer l'estimateur du quantile à l'étape $(n + 1)$ en fonction de celui à l'étape $n \mapsto$ difficulté à le faire pour les estimateurs 1 et 2 \mapsto Hypothèse simplificatrice : on considère le quantile empirique.

Cas où $[(n + 1)\alpha] = [n\alpha] + 1$ et où l'estimateur à l'étape n n'est pas la plus grande observation :



Une première simplification

Notre critère était

$$\Gamma_{n+1}(x_{n+1}) = \int_{\mathbb{X}} |\mathbb{P}_n(Y(x) \geq \hat{q}_\alpha^{n+1} | Y(x_{n+1}) = Y_{n+1}) - (1 - \alpha)| dx.$$

Une première simplification

Notre critère était

$$\Gamma_{n+1}(x_{n+1}) = \int_{\mathbb{X}} |\mathbb{P}_n(Y(x) \geq \hat{q}_\alpha^{n+1} | Y(x_{n+1}) = Y_{n+1}) - (1 - \alpha)| dx.$$

Nous pouvons aussi l'écrire :

$$\Gamma_{n+1}(x_{n+1}) = \int_{\mathbb{X}} \left| \int_{\mathbb{R}} \mathbb{P}_n(Y(x) \geq \hat{q}_\alpha^{n+1} | y_{n+1}) d\phi(y_{n+1}) - (1 - \alpha) \right| dx$$

où ϕ est la loi de $Y_n(x_{n+1})$ c'est à dire une $\mathcal{N}(m_n(x_{n+1}), s_n(x_{n+1}))$.

Si on lui applique la formule de mise à jour il devient (cas standard) :

$$\begin{aligned}
 P_n = & \int_{-\infty}^{y_{([n\alpha]+1)}} \mathbb{P}_n \left(Y(x) \geq \hat{q}_\alpha^n | y_{n+1} \right) d\phi(y_{n+1}) + \int_{y_{([n\alpha]+1)}}^{y_{([n\alpha]+2)}} \mathbb{P}_n \left(Y(x) \geq y_{n+1} | y_{n+1} \right) d\phi(y_{n+1}) \\
 & + \int_{y_{([n\alpha]+2)}}^{+\infty} \mathbb{P}_n \left(Y(x) \geq y_{([n\alpha]+2)} | y_{n+1} \right) d\phi(y_{n+1})
 \end{aligned}$$

Si on lui applique la formule de mise à jour il devient (cas standard) :

$$P_n = \int_{-\infty}^{y_{([n\alpha]+1)}} \mathbb{P}_n(Y(x) \geq \hat{q}_\alpha^n | y_{n+1}) d\phi(y_{n+1}) + \int_{y_{([n\alpha]+1)}}^{y_{([n\alpha]+2)}} \mathbb{P}_n(Y(x) \geq y_{n+1} | y_{n+1}) d\phi(y_{n+1}) \\ + \int_{y_{([n\alpha]+2)}}^{+\infty} \mathbb{P}_n(Y(x) \geq y_{([n\alpha]+2)} | y_{n+1}) d\phi(y_{n+1})$$

On voit donc qu'on doit exprimer des expressions de la forme

$$P_n^1 = \mathbb{P}_n(Y(x) \leq a | Y(x_{n+1}) \leq b) \text{ ou}$$

$$P_n^2 = \mathbb{P}_n(Y(x) \leq Y(x_{n+1}) | Y(x_{n+1}) \leq c).$$

Remarque : dépendance en y_{n+1}

$$P_n^1 = \int_{-\infty}^b \mathbb{P}_n(Y(x) \leq a | y_{n+1}) d\phi(y_{n+1}) = \int_{-\infty}^b \mathbb{P}_n\left(\frac{Y(x) - m_{n+1}(x)}{s_{n+1}(x)} \leq \frac{a - m_{n+1}(x)}{s_{n+1}(x)} | y_{n+1}\right) d\phi(y_{n+1})$$

Attention : cela dépend de y_{n+1} que nous ne connaissons pas \mapsto
 Utilisation des formules de mise à jour pour exprimer en fonction
 de la tribu \mathcal{F}_n .

Remarque : dépendance en y_{n+1}

$$P_n^1 = \int_{-\infty}^b \mathbb{P}_n(Y(x) \leq a | y_{n+1}) d\phi(y_{n+1}) = \int_{-\infty}^b \mathbb{P}_n\left(\frac{Y(x) - m_{n+1}(x)}{s_{n+1}(x)} \leq \frac{a - m_{n+1}(x)}{s_{n+1}(x)} | y_{n+1}\right) d\phi(y_{n+1})$$

Attention : cela dépend de y_{n+1} que nous ne connaissons pas \mapsto
 Utilisation des formules de mise à jour pour exprimer en fonction
 de la tribu \mathcal{F}_n .

En utilisant : $m_{n+1}(x) = m_n(x) + \frac{c_n(x, x_{n+1})(Y(x_{n+1}) - m_n(x_{n+1}))}{s_n^2(x_{n+1})}$, on
 enlève cette dépendance :

$$\int_{-\infty}^b \mathbb{P}_n\left(\frac{Y(x) - m_n(x)}{s_{n+1}(x)} \leq \frac{a - m_n(x)}{s_{n+1}(x)} | y_{n+1}\right) d\phi(y_{n+1})$$

En notant $f_i^j(d) = \frac{d - m_i(j)}{s_i(j)}$, nous avons donc :

$$P_n^1 = \mathbb{P}_n \left(\frac{Y(x) - m_n(x)}{s_{n+1}(x)} \leq f_n^x(a) \mid \frac{Y(x_{n+1}) - m_n(x_{n+1})}{s_n(x_{n+1})} \leq f_n^{x_{n+1}}(b) \right).$$

En notant $f_i^j(d) = \frac{d - m_i(j)}{s_i(j)}$, nous avons donc :

$$P_n^1 = \mathbb{P}_n \left(\frac{Y(x) - m_n(x)}{s_{n+1}(x)} \leq f_n^x(a) \mid \frac{Y(x_{n+1}) - m_n(x_{n+1})}{s_n(x_{n+1})} \leq f_n^{x_{n+1}}(b) \right).$$

Ainsi :

$$P_n^1 = \int_{-\infty}^{f_n^{x_{n+1}}(b)} \mathbb{P}_n \left(Y \leq f_n^x(a) \mid X = u \right) \frac{\exp\left(-\frac{u^2}{2}\right)}{\sqrt{2\pi}} du$$

où sachant la tribu \mathcal{F}_n , $X = \frac{Y(x_{n+1}) - m_n(x_{n+1})}{s_n(x_{n+1})} \sim \mathcal{N}(0, 1)$ et

$$Y = \frac{Y(x) - m_n(x)}{s_{n+1}(x)} \sim \mathcal{N}\left(0, \frac{s_n^2(x)}{s_{n+1}^2(x)}\right).$$

Propriétés sur les vecteurs gaussiens

Nous avons (X, Y) vecteur gaussien centrée de matrice de

Covariance $K_{(X,Y)} = K_{\beta} = \begin{pmatrix} 1 & \beta \\ \beta & 1 + \beta^2 \end{pmatrix}$, où l'on note

$$\beta = \frac{c_n(x, x_{n+1})}{s_{n+1}(x)s_n(x_{n+1})}.$$

Ainsi, la loi de Y sachant $X = x$ est une loi normale de paramètres

$$m_X = \frac{\text{Cov}(X, Y)_X}{\text{var}(X)} = \beta x \text{ et } \sigma^2 = \text{Var}(Y) - \frac{\text{Cov}(X, Y)^2}{\text{Var}(X)^2} = 1.$$

Finalement : $\mathcal{L}(Y|X = x) = \mathcal{N}(\beta x, 1)$.

Forme analytique

$$P_n^1 = \int_{-\infty}^{f_n^{x_{n+1}}(b)} \int_{-\infty}^{f_n^x(a)} \exp\left(-\frac{1}{2}((y - \beta u)^2 + u^2)\right) \frac{dy du}{2\pi}$$

Forme analytique

$$P_n^1 = \int_{-\infty}^{f_n^{x_{n+1}}(b)} \int_{-\infty}^{f_n^x(a)} \exp\left(-\frac{1}{2}((y - \beta u)^2 + u^2)\right) \frac{dy du}{2\pi}$$

Et par par Fubini :

$$\begin{aligned} &= \int_{]-\infty, f_n^{x_{n+1}}(b)] \times]-\infty, f_n^x(a)]} \exp\left(-\frac{1}{2}(u, y) K_\beta^{-1}(u, y)^T\right) \frac{d(u, y)}{2\pi} \\ &= \Phi_{K_\beta}\left(f_n^{x_{n+1}}(b), f_n^x(a)\right) \end{aligned}$$

Où Φ_K est la fonction de répartition d'un vecteur gaussien centré de matrice de covariance K .

$$\text{Simplification de } P_n^2 = \mathbb{P}_n \left(Y(x) \leq Y(x_{n+1}) \mid Y(x_{n+1}) \leq b \right)$$

Idée : enlever le $Y(x_{n+1})$ de la partie de gauche pour se ramener à une probabilité de la forme de la précédente.

Simplification de $P_n^2 = \mathbb{P}_n \left(Y(x) \leq Y(x_{n+1}) \mid Y(x_{n+1}) \leq b \right)$

Idée : enlever le $Y(x_{n+1})$ de la partie de gauche pour se ramener à une probabilité de la forme de la précédente.

$$X = \frac{Y(x_{n+1}) - m_n(x_{n+1})}{s_n(x_{n+1})} \sim \mathcal{N}(0, 1),$$

Simplification de $P_n^2 = \mathbb{P}_n \left(Y(x) \leq Y(x_{n+1}) \mid Y(x_{n+1}) \leq b \right)$

Idée : enlever le $Y(x_{n+1})$ de la partie de gauche pour se ramener à une probabilité de la forme de la précédente.

$$\begin{aligned}
 X &= \frac{Y(x_{n+1}) - m_n(x_{n+1})}{s_n(x_{n+1})} \sim \mathcal{N}(0, 1), \\
 Y' &= \frac{Y(x) - m_{n+1}(x)}{s_{n+1}(x)} - \frac{Y(x_{n+1}) - m_n(x_{n+1})}{s_{n+1}(x)} \\
 &\sim \mathcal{N}\left(0, \frac{s_n^2(x)}{s_{n+1}^2(x)} + \frac{s_n^2(x_{n+1})}{s_{n+1}^2(x)} - 2 \frac{c_n(x, x_{n+1})}{s_{n+1}^2(x)}\right)
 \end{aligned}$$

Simplification de $P_n^2 = \mathbb{P}_n \left(Y(x) \leq Y(x_{n+1}) \mid Y(x_{n+1}) \leq b \right)$

Idée : enlever le $Y(x_{n+1})$ de la partie de gauche pour se ramener à une probabilité de la forme de la précédente.

$$X = \frac{Y(x_{n+1}) - m_n(x_{n+1})}{s_n(x_{n+1})} \sim \mathcal{N}(0, 1),$$

$$Y' = \frac{Y(x) - m_{n+1}(x)}{s_{n+1}(x)} - \frac{Y(x_{n+1}) - m_n(x_{n+1})}{s_{n+1}(x)}$$

$$\sim \mathcal{N}\left(0, \frac{s_n^2(x)}{s_{n+1}^2(x)} + \frac{s_n^2(x_{n+1})}{s_{n+1}^2(x)} - 2 \frac{c_n(x, x_{n+1})}{s_{n+1}^2(x)}\right)$$

$$\text{Cov}(X, Y') = \frac{c_n(x, x_{n+1}) - s_n^2(x_{n+1})}{s_{n+1}(x)s_n(x_{n+1})} \text{ que l'on appelle } \nu.$$

Simplification de $P_n^2 = \mathbb{P}_n \left(Y(x) \leq Y(x_{n+1}) \mid Y(x_{n+1}) \leq b \right)$

Idée : enlever le $Y(x_{n+1})$ de la partie de gauche pour se ramener à une probabilité de la forme de la précédente.

$$X = \frac{Y(x_{n+1}) - m_n(x_{n+1})}{s_n(x_{n+1})} \sim \mathcal{N}(0, 1),$$

$$Y' = \frac{Y(x) - m_{n+1}(x)}{s_{n+1}(x)} - \frac{Y(x_{n+1}) - m_n(x_{n+1})}{s_{n+1}(x)}$$

$$\sim \mathcal{N}\left(0, \frac{s_n^2(x)}{s_{n+1}^2(x)} + \frac{s_n^2(x_{n+1})}{s_{n+1}^2(x)} - 2 \frac{c_n(x, x_{n+1})}{s_{n+1}^2(x)}\right)$$

$$\text{Cov}(X, Y') = \frac{c_n(x, x_{n+1}) - s_n^2(x_{n+1})}{s_{n+1}(x)s_n(x_{n+1})} \text{ que l'on appelle } \nu.$$

Nous avons alors avec cette notation $\text{Var}(Y') = 1 + \nu^2$ et

$$K_{(X, Y')} = K_\nu = \begin{pmatrix} 1 & \nu \\ \nu & 1 + \nu^2 \end{pmatrix}.$$

Et alors en utilisant la formule de mise à jour de l'espérance exactement comme pour P_n^1 :

$$P_n^2 = \mathbb{P}_n \left(Y' \leq z_n(x) \mid X \leq f_n^{x_{n+1}}(b) \right)$$

avec $z_n(x) = \frac{m_n(x_{n+1}) - m_n(x)}{s_{n+1}(x)}$ (qui ne dépend plus de $Y(x_{n+1})$).

Et alors en utilisant la formule de mise à jour de l'espérance exactement comme pour P_n^1 :

$$P_n^2 = \mathbb{P}_n \left(Y' \leq z_n(x) \mid X \leq f_n^{x_{n+1}}(b) \right)$$

avec $z_n(x) = \frac{m_n(x_{n+1}) - m_n(x)}{s_{n+1}(x)}$ (qui ne dépend plus de $Y(x_{n+1})$).

Nous sommes finalement ramené à simplifier une probabilité de la forme de P_n^1 avec X et Y' . En suivant exactement le même chemin que dans la partie précédente, nous trouvons alors que :

$$P_n^2 = \Phi_{K_\nu} \left(f_n^{x_{n+1}}, z_n(x) \right)$$

Critère simplifié dans le cas standard

Finalement, nous trouvons que si nous ne sommes pas dans un cas limite et que $[(n+1)\alpha] = [n\alpha] + 1$ alors :

$$\begin{aligned} \Gamma_{n+1}(x_{n+1}) = & \int_{\mathbb{X}} \left| \Phi_{K_{-\beta}} \left(f_n^{x_{n+1}}(y_{([n\alpha]+1)}), f_n^x(-\hat{q}_\alpha^n) \right) \right. \\ & + \Phi_{K_{-nu}} \left(f_n^{x_{n+1}}(y_{([n\alpha]+2)}), -z_n(x) \right) \\ & - \Phi_{K_{-\nu}} \left(f_n^{x_{n+1}}(y_{([n\alpha]+1)}), -z_n(x) \right) \\ & \left. + \Phi_{K_\beta} \left(-f_n^{x_{n+1}}(y_{([n\alpha]+2)}), -f_n^x(y_{([n\alpha]+2)}) \right) - (1 - \alpha) \right| dx \end{aligned}$$

Plus de dépendance en y_{n+1} , pas besoin de simuler de gaussienne pour l'estimer : gain de temps de calcul.

Fonction de répartition calculable directement avec le logiciel R : pas d'approximation MC.

On a une seule boucle par évaluation du critère en un x : celle pour estimer l'intégrale.

- 1 Modélisation du problème
 - Modèle
 - Rappels sur le krigeage
- 2 Choix des estimateurs
- 3 Stratégie de choix du design : la méthode SUR
 - Vers une méthode d'apprentissage actif
 - Sélection du nouveau point à l'étape n : méthode SUR
 - Algorithme
- 4 Une autre méthode venant de l'optimisation bayésienne
 - Idée
 - A la recherche d'une forme analytique de notre critère
 - Algorithme
- 5 Résultats
 - En dimension 2
 - Dimension 5

Algorithme

1. On tire une grille $x_{\text{suiv}} (\mathcal{X})$ de 500 points sur $[0, 1]^d$.
2. On prend les n_0 premiers points de \mathcal{X} pour \mathcal{X}_0 . On évalue ses points et on met les évaluations dans un vecteur y , on évalue le quantile courant.
3. On tire une grille x_{grid} de points de $[0, 1]^d$ qui nous permettra d'estimer l'intégrale sur \mathbb{X} .
4. On appelle la fonction krigeage pour récupérer tous les coefficients dont nous aurons besoin pour initialiser les formules de MAJ.

Algorithme suite

6. Pour n allant de n_0 à n_{budget} :
 - a. On passe dans la fonction "nextPoint" qui rend le nouveau point à évaluer.
 - b. On ajoute le nouveau point à x , on l'évalue, on ajoute le résultat à y et on calcule l'estimateur courant.
 - c. On utilise les formules de mise à jour et on stocke les nouveaux paramètres nécessaires pour la prochaine étape.
 - c. On enlève le point choisi de la grille x_{suiv} .
5. On renvoie les deux estimateurs.

Détail de la fonction nextPoint

Entrées : `xgrid`, `x`, `y`, `xsuiv` courants et toutes les variables stockées à l'étape précédente nécessaires à cette étape.

- I. Récupération des variables de stock et création de nouvelles variables de stock.
- II. Boucle sur les `xsuiv` (dans le but de remplir un vecteur intégrale qui nous donne en coordonnée `i`, l'approximation de l'intégrale si x_{n+1}^* est le `xsuiv` courant).
 - A. Calcul par formules de mise à jour des coefficients utiles au calcul des fonctions de répartitions, ne dépendant que de `xsuiv`. Stockage.
 - B. Boucle sur les `xgrid` (dans le but de remplir un vecteur `contrib` qui aura en coordonnée `k` la contribution de l'intégrale donnée par le `xgrid` suivant en considérant que x_{n+1}^* est le `xsuiv` courant).

Détails sur "nextpoint" suite

- a. Calcul par formules de mise à jours des coefficients utiles au calcul des fonctions de répartition bi-gaussiennes. Stockage
 - b. Calculs des fonctions de répartition bi-gaussiennes avec le package bivnorm de r (en distinguant les deux cas sur la partie entière de $n\alpha$).
 - c. Calcul de la contribution courante (en distinguant si l'estimateur courant et la plus petite, la plus grande ou une autre observation).
- C. Calcul de l'approximation de l'intégrale courante grâce au vecteur des contributions.
- II. Recherche du x_{suiv} qui minimise le vecteur intégrale.

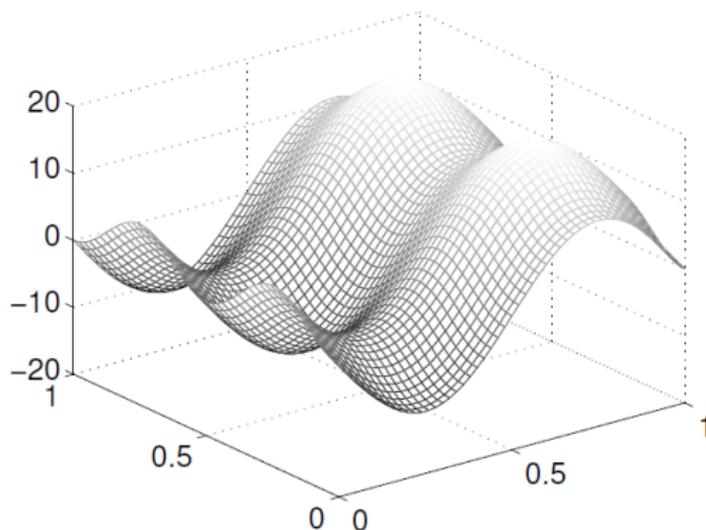
Sortie : Le nouveau point à évaluer et les variables stockées. 

- 1 Modélisation du problème
 - Modèle
 - Rappels sur le krigeage
- 2 Choix des estimateurs
- 3 Stratégie de choix du design : la méthode SUR
 - Vers une méthode d'apprentissage actif
 - Sélection du nouveau point à l'étape n : méthode SUR
 - Algorithme
- 4 Une autre méthode venant de l'optimisation bayésienne
 - Idée
 - A la recherche d'une forme analytique de notre critère
 - Algorithme
- 5 Résultats
 - En dimension 2
 - Dimension 5

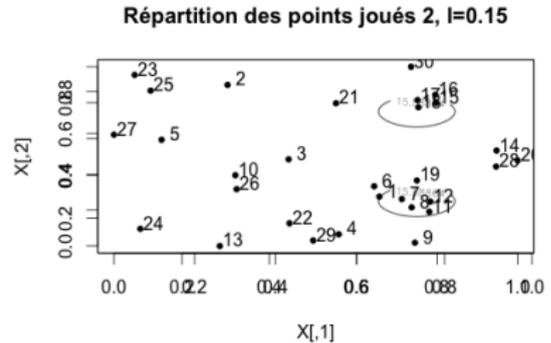
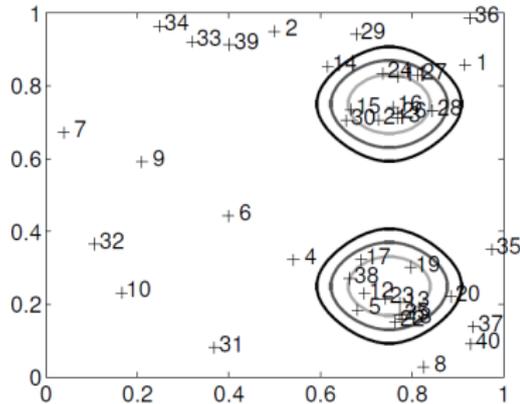
Résultats en dimension 2

- $n_{suiv} = 500$
- MC1 et MC2 = 10
- $n_{grid} = 100$
- $n_0 = 10, 20$
- $n_{budget} = 40, 70, 100$
- $\alpha = 0.95$
- noyau gaussien : $k(x, y) = \exp\left(\frac{-\|x-y\|^2}{2l^2}\right)$ avec $l = 0.15$

Fonction de Ishigami : $f(x, y) = \sin(\pi(2x - 1)) + 7\sin(\pi(2y - 1))^2 + 0.1\pi^4\sin(\pi(2x - 1))$



Points joués pour la fonction Ishigami



Erreur relative par rapport au quantile estimé avec les 500 points de la grille x_{suiv} par la méthode SUR

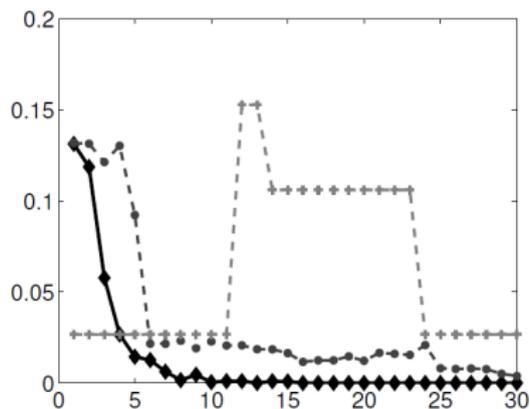
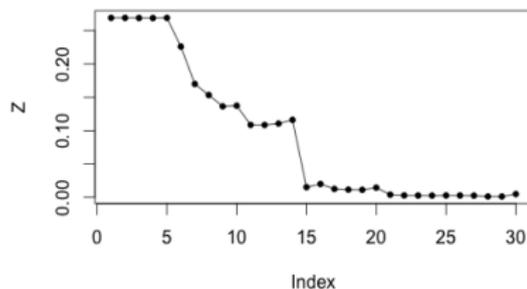


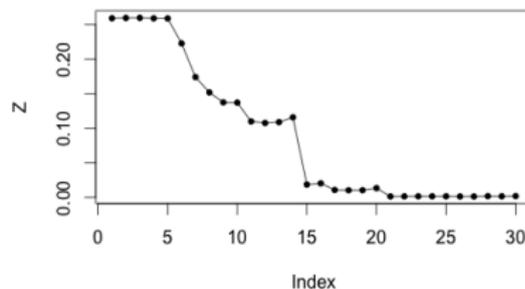
FIGURE: En diamant noir en ligne pleine : SUR

Erreur relative par rapport au quantile estimé avec les 500 points de x_{suiv} par la nouvelle méthode avec estimateur 1 puis estimateur 2

Erreur relative est1 (500) , $l=0.15$

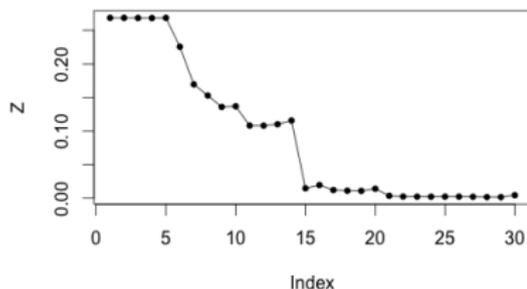


Erreur relative est2 (500), $l=0.15$

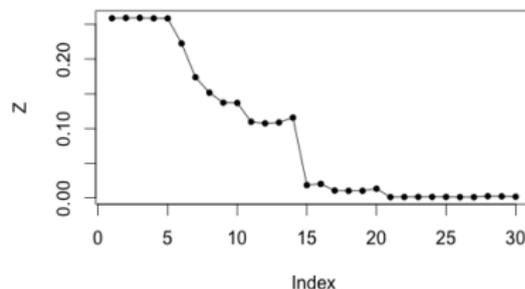


Erreur relative par rapport au vrai quantile par la méthode 2 avec estimateur 1 puis estimateur 2

Erreur relative est1 (10^5), $l=0.15$

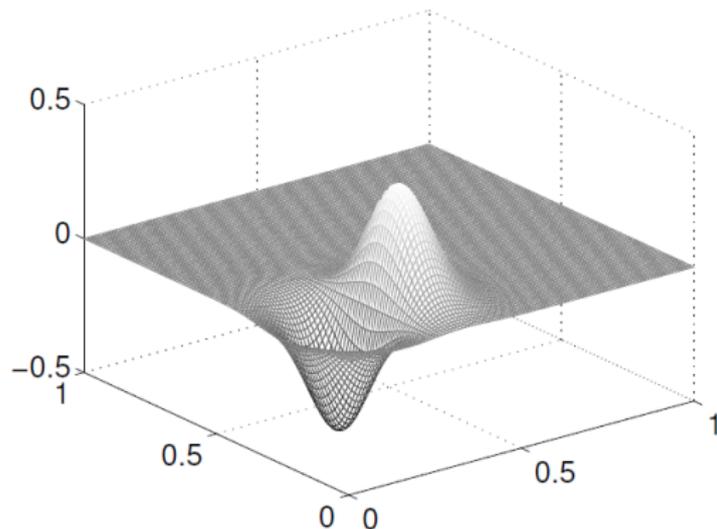


Erreur relative est2 (10^5), $l=0.15$

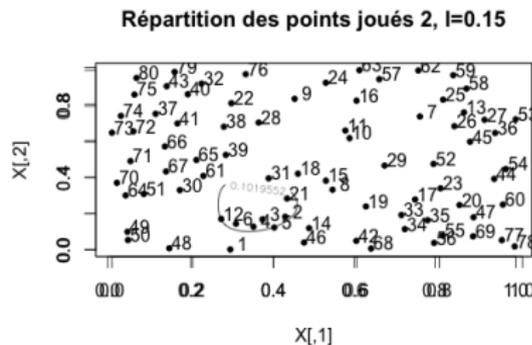
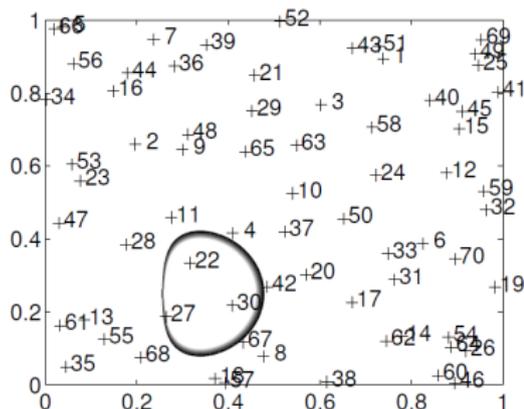


Fonction de Gramacy :

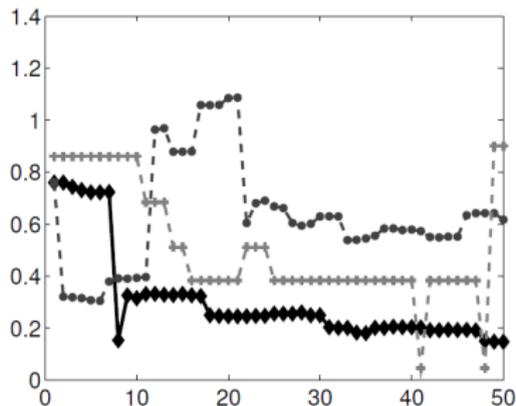
$$f(v, w) = (8v - 2) \exp(-(8v - 2)^2 - (8w - 2)^2)$$



Points joués pour la fonction Gramacy

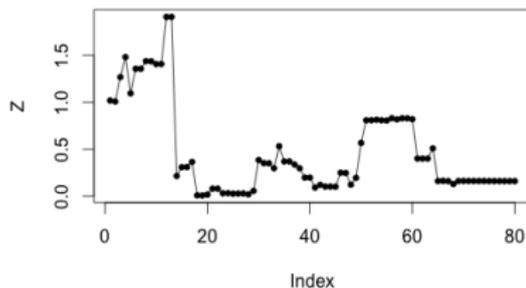


Erreur relative par rapport au quantile estimé avec les 500 points de la grille par la méthode SUR

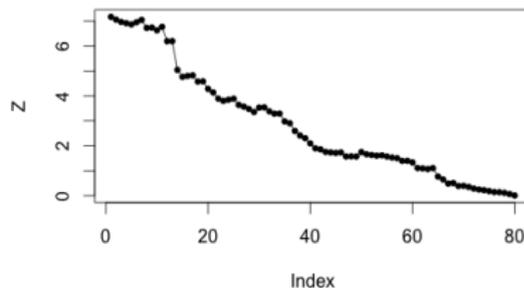


Erreur relative par rapport au quantile estimé avec 500 points par la nouvelle méthode avec estimateur 1 puis estimateur 2

Erreur relative est1 (500) , $l=0.15$

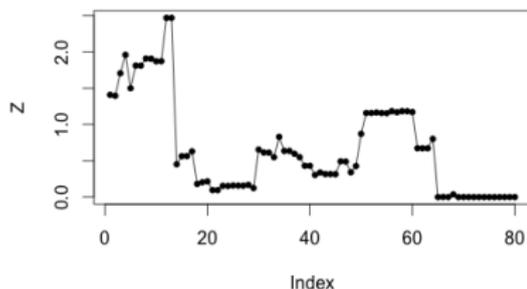


Erreur relative est2 (500), $l=0.15$

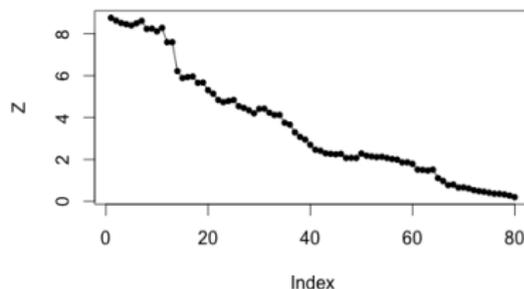


Erreur relative par rapport au vrai quantile par la nouvelle méthode avec estimateur 1 puis estimateur 2

Erreur relative est1 (10^5), $l=0.15$



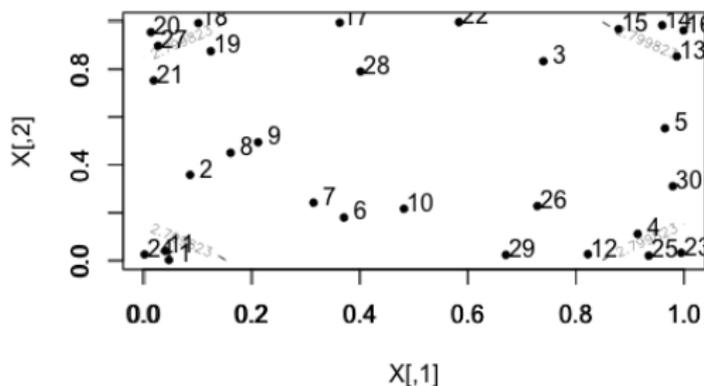
Erreur relative est2 (10^5), $l=0.15$



Fonction de Sobol en dimension 2 :

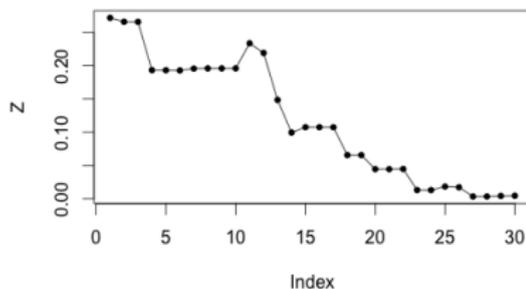
$$f(x, y) = |4x - 2| \times |4y - 2|$$

Répartition des points joués 2, $l=0.15$

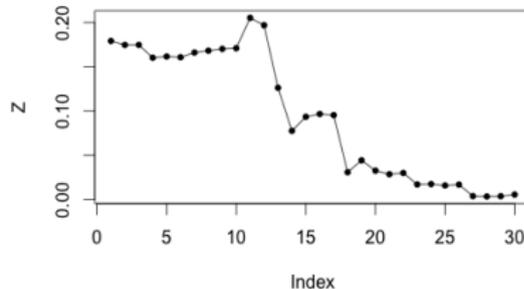


Erreur relative par rapport au quantile estimé avec 500 points par la nouvelle méthode avec estimateur 1 puis estimateur 2

Erreur relative est1 (500) , $l=0.15$

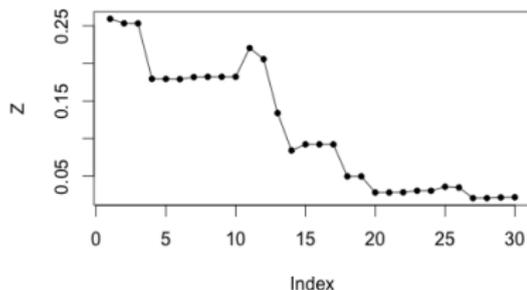


Erreur relative est2 (500), $l=0.15$

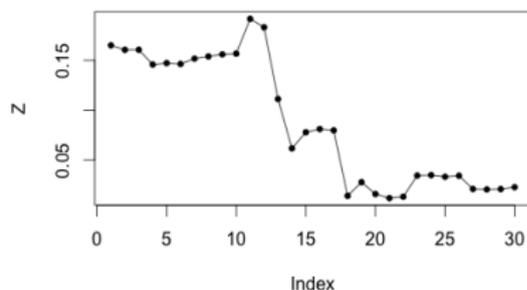


Erreur relative par rapport au vrai quantile par la nouvelle méthode avec estimateur 1 puis estimateur 2

Erreur relative est1 (10^5), $l=0.15$



Erreur relative est2 (10^5), $l=0.15$



SUR :

- Ishigami (10+30) : 4h00
- Gramacy (20+50) : 5h30

Nouvelle méthode

- Ishigami (10+30) : 40 min
- Gramacy (20+ 50) : 1h30, (20+ 80) 2h15

- 1 Modélisation du problème
 - Modèle
 - Rappels sur le krigeage
- 2 Choix des estimateurs
- 3 Stratégie de choix du design : la méthode SUR
 - Vers une méthode d'apprentissage actif
 - Sélection du nouveau point à l'étape n : méthode SUR
 - Algorithme
- 4 Une autre méthode venant de l'optimisation bayésienne
 - Idée
 - A la recherche d'une forme analytique de notre critère
 - Algorithme
- 5 Résultats
 - En dimension 2
 - Dimension 5

SUR dimension 5 : trop grosse complexité.

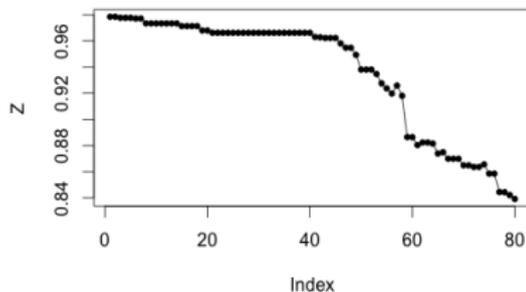
SUR dimension 5 : trop grosse complexité.

Nouvelle méthode : Avec les mêmes paramètres que la dimension 2 pour Sobol :

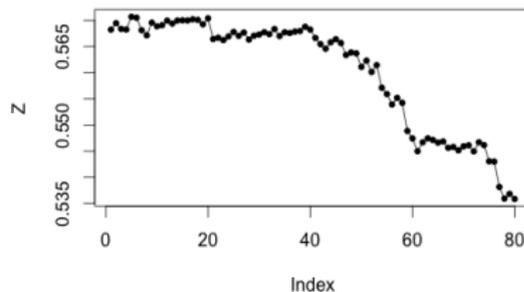
- 8h de calcul
- Mauvaise précision car grilles trop petites

Erreur relative par rapport au quantile estimé avec 500 points par la nouvelle méthode avec estimateur 1 puis estimateur 2

Erreur relative est1 (500) , $l=0.15$

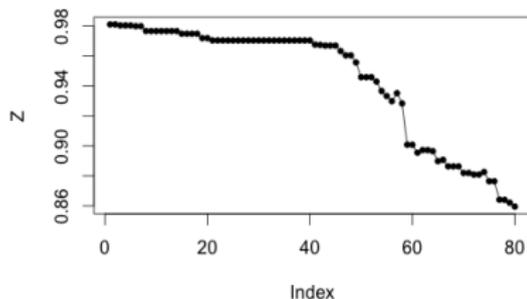


Erreur relative est2 (500), $l=0.15$

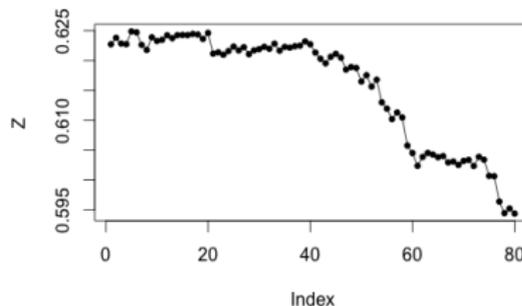


Erreur relative par rapport au vrai quantile par la méthode 2 avec estimateur 1 puis estimateur 2

Erreur relative est1 (10^5), $l=0.15$



Erreur relative est2 (10^5), $l=0.15$



Conclusion et perspectives

Nouvel algorithme de complexité de calcul très réduite, espoir d'aller jusqu'à la dimension 5.

Conclusion et perspectives

Nouvel algorithme de complexité de calcul très réduite, espoir d'aller jusqu'à la dimension 5.

Perspectives :

- Trouver une formule de mise à jour de l'estimateur du quantile avec un bon estimateur.
- Chercher un moyen plus intelligent d'atteindre le minimum (algorithme stochastique, dichotomie...).