

Méthodes statistiques et d'apprentissage pour l'estimation de quantiles et de superquantiles dans des modèles de codes numériques ou stochastiques

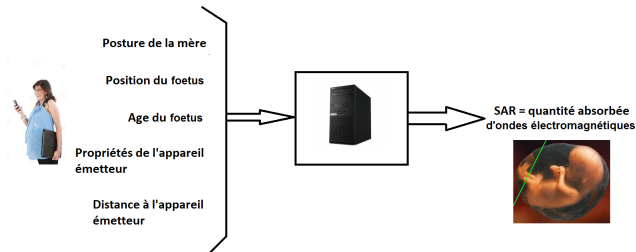
Tatiana Labopin-Richard

Soutenance de thèse - EDMITT

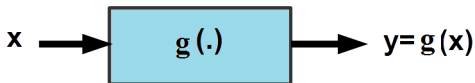
1er juillet 2016

Qu'est-ce qu'un code numérique ?

Question : quelle quantité d'ondes électromagnétiques absorbe le fœtus d'une femme enceinte utilisant un téléphone portable ?



Modélisation d'un code numérique



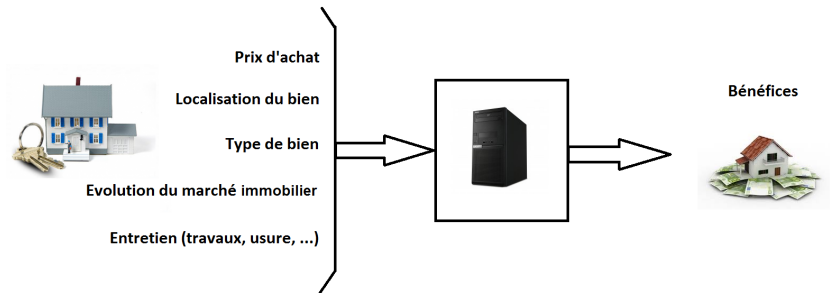
- Un **code numérique** est une fonction $g : \mathbb{X} \subset \mathbb{R}^d \longrightarrow \mathbb{R}$.
- **Incertitude** sur les entrées $\implies X$ est **aléatoire**.
- La sortie du code numérique est $Y = g(X)$.

Inférence sur la sortie d'un code numérique

On cherche des informations sur la loi de la sortie $Y = g(X)$.

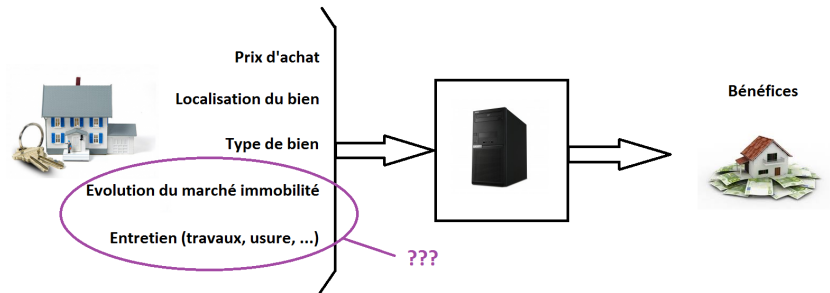
Qu'est-ce qu'un code stochastique ?

Question : quel bénéfice tire-t-on lors d'un achat immobilier ?



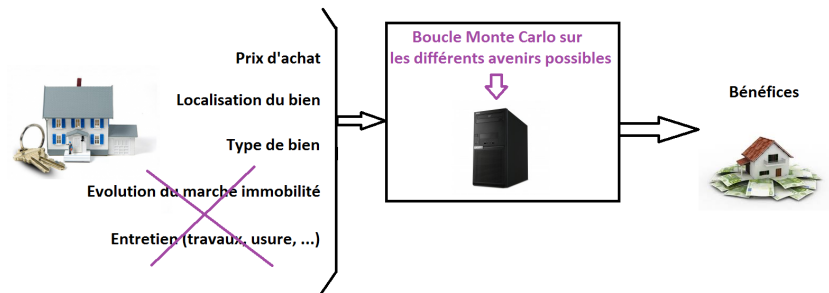
Qu'est-ce qu'un code stochastique ?

Question : quel bénéfice tire-t-on lors d'un achat immobilier ?

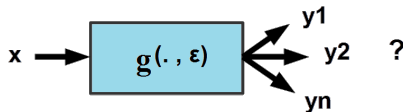


Qu'est-ce qu'un code stochastique ?

Question : quel bénéfice tire-t-on lors d'un achat immobilier ?



Modélisation d'un code stochastique



- Un **code stochastique** est une fonction $g : \mathbb{X} \subset \mathbb{R}^d \times \mathbb{R}^l \rightarrow \mathbb{R}$.
- Les entrées X sont aléatoires (incertitudes).
- La graine aléatoire ϵ modélise la **stochasticité interne** au code.
- La sortie du code stochastique est $Y = g(X, \epsilon)$.

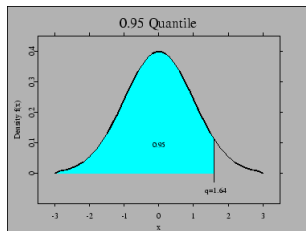
Inférence sur la sortie d'un code stochastique

On cherche à avoir des informations sur la loi de $Y(x) = g(x, \epsilon)$, pour tout $x \in \mathbb{X} \subset \mathbb{R}^d$.

Quantile

On étudie des codes dont on veut savoir **contrôler la sortie**. On cherche donc à estimer des **mesures de risque** $\mathcal{R} : \mathcal{M}(\mathbb{R}) \rightarrow \mathbb{R}$.

- Le **quantile** de niveau $\alpha \in]0, 1[$, $q_\alpha(Z) := F_Z^{-1}(\alpha)$.



Problèmes : le quantile ...

- 1) ne donne pas assez d'information sur la queue de distribution.
- 2) n'est pas sous-additif (parfois $q_\alpha(Z + Z') > q_\alpha(Z) + q_\alpha(Z')$).

Mesure de risque cohérente

Définition - Rockafellar (2007)

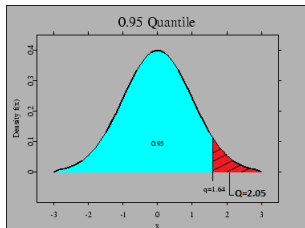
Soit $\mathcal{R}$ une mesure de risque et Z et Z' deux variables aléatoires réelles. On dit que $\mathcal{R}$ est cohérente si, et seulement si, elle satisfait les cinq axiomes suivants :

- i) **Invariance par constante** : soit $C \in \mathbb{R}$, si $X = C$ (p.s.) alors $\mathcal{R}(C) = C$.
- ii) **Homogénéité** : $\forall \lambda > 0$, $\mathcal{R}(\lambda Z) = \lambda \mathcal{R}(Z)$.
- iii) **Sous-additivité** : $\mathcal{R}(Z + Z') \leq \mathcal{R}(Z) + \mathcal{R}(Z')$.
- iv) **Croissance** : si $Z \leq Z'$ (p.s.) alors $\mathcal{R}(Z) \leq \mathcal{R}(Z')$.
- v) **Fermeture** : soit $(Z_h)_{h \in \mathbb{R}}$ une suite de variables aléatoires. Si $\mathcal{R}(Z_h) \leq 0$ et $\lim_{h \rightarrow 0} \|Z_h - Z\|_2 = 0$ alors $\mathcal{R}(Z) \leq 0$.

Le superquantile - Rockafellar (2007)

- Le **superquantile** de niveau α ,

$$Q_\alpha(Z) := \mathbb{E}(Z | Z \geq q_\alpha(Z)) = \mathbb{E} \left(\frac{Z \mathbf{1}_{Z \geq F_Z^{-1}(\alpha)}}{1 - \alpha} \right)$$



Propriété (Rockafellar - 2007)

Lorsque la fonction de répartition F_Z est continue, le superquantile est une mesure de risque **cohérente**.

Bilan

Problématique

Comment estimer les deux **mesures de risque**

$$q_\alpha = F^{-1}(\alpha) \text{ et } Q_\alpha = \frac{1}{1-\alpha} \mathbb{E}(Y | Y \geq q_\alpha),$$

pour les **sorties de codes numériques ou stochastiques**

$$Y = g(X) \text{ ou } Y(x) = g(x, \epsilon), \forall x \in \mathbb{X} ?$$

- Besoin de méthodes différentes suivant le **type de code** et le **temps de calcul** pour obtenir une évaluation du code.

- 1 Estimation empirique
 - Le superquantile
 - Une nouvelle mesure de risque : le superquantile de Bregman
- 2 Estimation séquentielle par algorithme stochastique et méthode des k plus proches voisins
 - Construction de l'algorithme
 - Résultats
 - Simulations numériques
- 3 Estimation séquentielle bayésienne et planification d'expériences
 - Etat de l'art
 - Deux nouvelles stratégies
 - Résultats numériques

1 Estimation empirique

- Le superquantile
- Une nouvelle mesure de risque : le superquantile de Bregman

2 Estimation séquentielle par algorithme stochastique et méthode des k plus proches voisins

- Construction de l'algorithme
- Résultats
- Simulations numériques

3 Estimation séquentielle bayésienne et planification d'expériences

- Etat de l'art
- Deux nouvelles stratégies
- Résultats numériques

Estimation empirique

CADRE : estimation de **superquantiles** en **un code numérique peu coûteux**.

Estimation empirique :

- 1) On tire un échantillon indépendant $(X_1, \dots, X_n)$ suivant la loi des entrées X .
- 2) On fournit chacune des entrées X_i au code.
- 3) On observe les réponses du code correspondant à chaque entrée : $Y_i = g(X_i)$.
- 4) On **calcule les quantités d'intérêt sur l'échantillon** $(Y_1, \dots, Y_n)$ de la loi Y .

1 Estimation empirique

- Le superquantile
- Une nouvelle mesure de risque : le superquantile de Bregman

2 Estimation séquentielle par algorithme stochastique et méthode des k plus proches voisins

- Construction de l'algorithme
- Résultats
- Simulations numériques

3 Estimation séquentielle bayésienne et planification d'expériences

- Etat de l'art
- Deux nouvelles stratégies
- Résultats numériques

Le cas du superquantile

- Estimateur empirique du quantile : $q_n := Y_{(\lfloor n\alpha \rfloor + 1)}$
- Estimateur empirique du superquantile :

$$Q_n := \frac{1}{(1 - \alpha)n} \sum_{i=\lfloor n\alpha \rfloor + 1}^n Y_{(i)} \quad (1)$$

Hypothèses - Comportement asymptotique de l'estimateur (1)

- H1)** La fonction quantile F_Y^{-1} est $\mathcal{C}^1$ sur $]0, 1[$ et sa dérivée satisfait $l = O((1 - t)^{-2 + \epsilon_l})$ pour un $\epsilon_l > 0$ au voisinage de 1^- .
- H2)** F_Y^{-1} est $\mathcal{C}^2$ sur $]0, 1[$ et sa dérivée seconde vérifie $L = O\left((1 - t)^{-\frac{5}{2} + \epsilon_L}\right)$ pour un $\epsilon_L > 0$ au voisinage de 1^- .

Comportement asymptotique de l'estimateur (1)

Proposition (L.-R. et al. - 2016)

- i) Sous **H1**, l'estimateur empirique (1) est consistant en probabilité

$$Q_n \xrightarrow[n \rightarrow +\infty]{\mathbb{P}} Q_\alpha.$$

- ii) Sous **H2**, l'estimateur empirique (1) est asymptotiquement gaussien

$$\sqrt{n}(Q_n - Q_\alpha) \Rightarrow \mathcal{N}\left(0, \frac{\sigma^2}{(1 - \alpha)^2}\right),$$

avec

$$\sigma^2 := \int_{\alpha}^1 \int_{\alpha}^1 \frac{(\min(x, y) - xy)}{f_Y(F_Y^{-1}(x))f(F_Y^{-1}(y))}$$

Exemple de la loi de Pareto

Pour $a > 0$, $F(t) = 1 - x^{-a}$, $F^{-1}(t) = (1 - t)^{-\frac{1}{a}}$

Le **superquantile est fini si et seulement si $a > 1$** et dans ce cas

$$Q_\alpha = (1 - \alpha)^{-\frac{1}{a}} \frac{a}{a-1}.$$

- $l(t) = (a(1 - t)^{-1-\frac{1}{a}})$, donc $l(t) = O((1 - t)^{-2+\epsilon_l})$ dès que $a > 1$. La **consistance est donc vraie pour $a > 1$** .
- $L(t) = C(a)(1 - x)^{-\frac{1}{a}-2}$ donc, $L(t) = O((1 - t)^{-\frac{5}{2}+\epsilon_L})$ dès que $a > 2$. **La normalité asymptotique est donc vraie pour $a > 2$** .

- 1 Estimation empirique
 - Le superquantile
 - Une nouvelle mesure de risque : le superquantile de Bregman
- 2 Estimation séquentielle par algorithme stochastique et méthode des k plus proches voisins
 - Construction de l'algorithme
 - Résultats
 - Simulations numériques
- 3 Estimation séquentielle bayésienne et planification d'expériences
 - Etat de l'art
 - Deux nouvelles stratégies
 - Résultats numériques

Vers une nouvelle mesure de risque

Définition : divergence de Bregman (Bregman - 1951)

Soit γ une **fonction strictement convexe**, de $\overline{\mathbb{R}}$ dans $\mathbb{R}$. On suppose que $\text{dom}\gamma$ est un ouvert non vide et que γ est propre, fermée et différentiable sur l'intérieur de $\text{dom}\gamma$.

La **divergence de Bregman** d_γ associée à γ est une fonction définie sur $\text{dom}\gamma \times \text{dom}\gamma$ par

$$d_\gamma(x, x') := \gamma(x) - \gamma(x') - \gamma'(x')(x - x') \quad ; (x, x' \in \text{dom}\gamma).$$

Vers une nouvelle mesure de risque

Définition : moyenne de Bregman (Bregman - 1967)

Soit μ une **mesure de probabilité** dont le support est inclus dans $\text{dom}\gamma$ telle que $\mu(\overline{\text{dom}\gamma} \setminus \text{dom}\gamma) = 0$ et γ' soit intégrable par rapport à μ . **La moyenne de Bregman est l'unique point** b dans le support de μ satisfaisant

$$\int d_\gamma(b, x) \mu(dx) = \min_{m \in \text{dom}\gamma} \int d_\gamma(m, x) \mu(dx). \quad (2)$$

En dérivant, on voit que

$$b = \gamma'^{-1} \left[\int \gamma'(x) \mu(dx) \right].$$

Vers une nouvelle mesure de risque

Définition : superquantile de Bregman (L.-R et al. - 2016)

Soit $\alpha \in]0, 1[$, le **superquantile de Bregman** $Q_\alpha^{d_\gamma}$ d'une variable aléatoire réelle Z est défini par

$$Q_\alpha^{d_\gamma} := \gamma'^{-1} \left(\mathbb{E}(\gamma'(Z) | Z \geq F_Z^{-1}(\alpha)) \right) = \gamma'^{-1} \left[\mathbb{E} \left(\frac{\gamma'(Z) \mathbf{1}_{Z \geq F_Z^{-1}(\alpha)}}{1 - \alpha} \right) \right].$$

En résumé, $Q_\alpha^{d_\gamma}$ satisfait (2) en prenant pour μ la distribution de Z conditionnellement à $Z \geq F_Z^{-1}(\alpha)$.

- Remarque :

$$Q_\alpha^{d_\gamma}(Z) = (\gamma')^{-1} [Q_\alpha(\gamma'(Z))].$$

Cohérence du Superquantile de Bregman

Proposition (L.-R et al. - 2016)

Soit α dans $]0, 1[$.

- i) Tout superquantile de Bregman satisfait les propriétés d'**invariance par constante** et de **croissance**.
- ii) Le superquantile de Bregman associé à γ est **homogène** si, et seulement si $\gamma''(x) = \beta x^\delta$ pour *deux réels* $\beta > 0$ et δ .
- iii) Si γ' est **concave et sous-additive**, alors les axiomes de **sous-additivité** et de **fermeture** sont vrais.

Estimateur empirique du superquantile de Bregman

- Estimateur empirique de superquantile de Bregman

$$\widehat{Q}_{\alpha}^{d_{\gamma}} = \gamma'^{-1} \left[\frac{1}{1-\alpha} \left(\frac{1}{n} \sum_{i=\lfloor n\alpha \rfloor + 1}^n \gamma'(Y_{(i)}) \right) \right], \quad (3)$$

Hypothèses pour l'étude du comportement asymptotique de l'estimateur (3)

- H3)** La fonction $\gamma' \circ F_Y^{-1}$ est $\mathcal{C}^1$ sur $]0, 1[$ et sa dérivée satisfait $l_{\gamma}(t) = O\left((1-t)^{-2+\epsilon_{l_{\gamma}}}\right)$ pour un $\epsilon_{l_{\gamma}} > 0$ au voisinage de 1^- .
- H4)** La fonction $\gamma' \circ F_Y^{-1}$ est $\mathcal{C}^2$ sur $]0, 1[$ et sa dérivée seconde satisfait $L_{\gamma}(t) = O\left((1-t)^{-\frac{5}{2}+\epsilon_{L_{\gamma}}}\right)$ pour un $\epsilon_{L_{\gamma}} > 0$ au voisinage de 1^- .

Comportement asymptotique de l'estimateur (3)

Proposition (L.-R. et al. - 2016)

i) Sous l'hypothèse **H3**, l'estimateur $\widehat{Q}_\alpha^{d_\gamma}$ est consistant :

$$\widehat{Q}_\alpha^{d_\gamma} \xrightarrow[n \rightarrow +\infty]{\mathbb{P}} Q_\alpha^{d_\gamma}.$$

ii) Sous l'hypothèse **H4**, l'estimateur $\widehat{Q}_\alpha^{d_\gamma}$ est normalement asymptotique : pour $Z = \gamma'(Y)$,

$$\sqrt{n} \left(\widehat{Q}_\alpha^{d_\gamma} - Q_\alpha^{d_\gamma} \right) \xrightarrow[n \rightarrow \infty]{\Rightarrow} \mathcal{N} \left(0, \frac{\sigma_\gamma^2}{\left(\gamma'' \left(Q_\alpha^{d_\gamma}(Y) \right) \right)^2 (1 - \alpha)^2} \right),$$

avec
$$\sigma_\gamma^2 := \int_\alpha^1 \int_\alpha^1 \frac{(\min(x, y) - xy)}{f_Z(F_Z^{-1}(x)) f_Z(F_Z^{-1}(y))} dx dy.$$

Exemple : loi de Pareto et fonction Bregman géométrique

Fonction de **Bregman géométrique** :

$$\gamma(x) = x \ln(x) - x + 1, \quad \gamma'(x) = \ln(x).$$

Ainsi, **pour tout** $a > 0$, $Q_{\alpha}^{d\gamma} = a^{-1} (1 - \ln(1 - \alpha))$ est **fini**.

- L'estimateur (3) est **consistant pour tout** $a > 0$ car

$$l_{\gamma}(t) = \frac{1}{a} \frac{1}{1-t} = O\left(\frac{1}{(1-t)^{2-\frac{1}{2}}}\right).$$

- L'estimateur (3) est **normalement asymptotique pour tout** $a > 0$ car

$$L_{\gamma}(t) = \frac{1}{a} \frac{1}{(1-t)^2} = O\left((1-t)^{-\frac{5}{2}+\frac{1}{3}}\right).$$

Simulations numériques pour la loi de Pareto

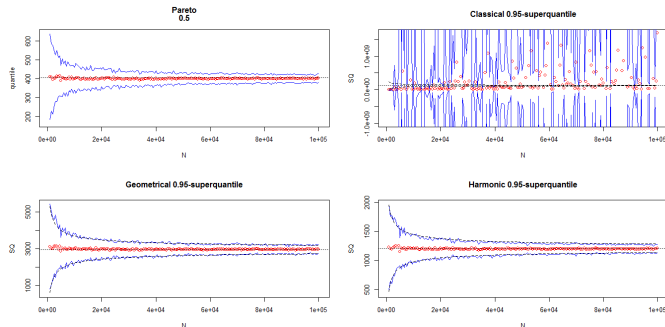


FIGURE – Test numérique de convergence pour la loi de Pareto ($a = 0.5$).

Simulations numériques pour la loi de Pareto

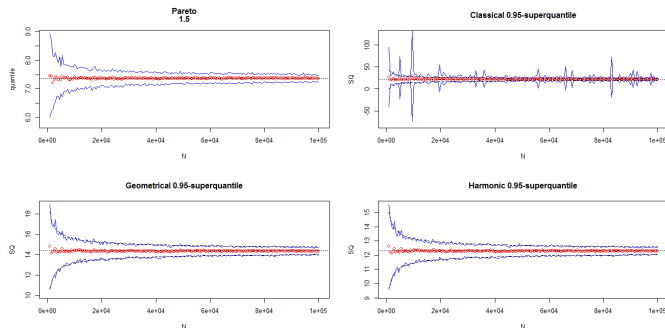


FIGURE – Test numérique de convergence pour la loi de Pareto ($a = 1.5$).

Simulations numériques pour la loi de Pareto

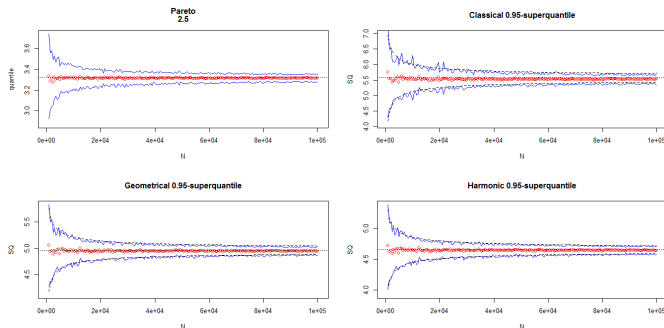


FIGURE – Test numérique de convergence pour la loi de Pareto ($a = 2.5$).

Application

Le logiciel GASCON a été créé par le CEA pour évaluer, à partir d'une source fictive, la quantité reçue d'ondes radioactives par une population exposée.

- Entrées : interactions hommes-nature, météo (vent, pluie, ...), distance à l'émission, etc...
- Sortie : dose effective reçue, par année et par personne adulte.

Quantile	Classical superquantile	Geometrical superquantile	Harmonic superquantile
1.304×10^{-13}	4.769×10^{-13}	3.316×10^{-13}	2.637×10^{-13}

Application

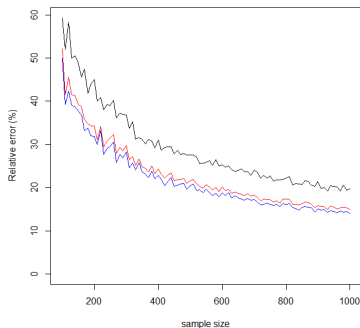


FIGURE – Erreur relative pour l'estimation du superquantile (noir), du superquantile géométrique (rouge) et du superquantile harmonique (bleu).

Bilan

- Nous avons exhibé des estimateurs empiriques ayant un **comportement asymptotique** intéressant pour estimer nos deux mesures de risques dans des codes numériques peu coûteux.
- Nous avons introduit une **nouvelle mesure de risque adaptée à un plus grand nombre de lois et dont l'estimation empirique est plus stable**.
- Les simulations numériques ainsi qu'un exemple applicatif illustrent la pertinence de l'introduction de cette nouvelle mesure de risque.

Perspectives :

- Mieux comprendre comment **choisir** la fonction γ suivant les problèmes posés.

- 1 Estimation empirique
 - Le superquantile
 - Une nouvelle mesure de risque : le superquantile de Bregman
- 2 Estimation séquentielle par algorithme stochastique et méthode des k plus proches voisins
 - Construction de l'algorithme
 - Résultats
 - Simulations numériques
- 3 Estimation séquentielle bayésienne et planification d'expériences
 - Etat de l'art
 - Deux nouvelles stratégies
 - Résultats numériques

Une première idée...

CADRE : estimation de **quantiles** en **code stochastique coûteux**.

Pour trouver le quantile de $Y(x) = g(x, \epsilon)$, on pourrait :

- 1) Fournir plusieurs fois la même entrée x au code stochastique pour construire un échantillon de la loi cible.
- 2) Utiliser l'estimateur empirique précédemment introduit.

Problèmes

- On veut estimer le quantile **pour toute entrée x** .
- Le code est **coûteux**, on ne peut donc pas forcément entrer dans le régime asymptotique.

Une méthode séquentielle pour un code coûteux

Une **méthode séquentielle à un pas** permet de gérer la précision en fonction du budget d'évaluations.

Algorithme stochastique pour estimer le quantile de la loi Z
(Robbins et Monro - 1951)

$$\begin{cases} \theta_0 \in \mathbb{R} \\ \theta_{n+1} = \theta_n - \frac{C}{n^\gamma} (\mathbf{1}_{Z_{n+1} \leq \theta_n} - \alpha) . \end{cases}$$

$\Rightarrow$ consistant et normalement gaussien si $\gamma \in]1/2, 1]$ et C bien choisi.

Etude d'un code stochastique

Stratégie adaptée au code stochastique

- 1) Un budget N d'appels au code est fixé.
- 2) On tire un échantillon $(X_1, \dots, X_N)$ de la loi des entrées X .
- 3) On observe les réponses correspondantes $(Y_1, \dots, Y_N)$.
- 4) **On applique un algorithme qui permet, pour tout x et n'utilisant que les précédentes observations, d'estimer le quantile conditionnel.**

Un algorithme stochastique plongé dans la méthode des k -plus proches voisins

L'algorithme

$$\begin{cases} \theta_0 & \in \mathbb{R} \\ \theta_{n+1} & = \theta_n - \frac{1}{n^\gamma} \left(\mathbf{1}_{Y(x)_{n+1} \leq \theta_n} - \alpha \right) \end{cases}$$

Un algorithme stochastique plongé dans la méthode des k -plus proches voisins

L'algorithme

$$\begin{cases} \theta_0 & \in \mathbb{R} \\ \theta_{n+1} & = \theta_n - \frac{1}{n^\gamma} \left(\mathbf{1}_{Y_{n+1} \leq \theta_n} - \alpha \right) \end{cases}$$

Un algorithme stochastique plongé dans la méthode des k -plus proches voisins

L'algorithme

$$\begin{cases} \theta_0(x) \in \mathbb{R} \\ \theta_{n+1}(x) = \theta_n(x) - \frac{1}{n^\gamma} \left(\mathbf{1}_{Y_{n+1} \leq \theta_n(x)} - \alpha \right) \end{cases}$$

Un algorithme stochastique plongé dans la méthode des k -plus proches voisins

L'algorithme

$$\begin{cases} \theta_0(x) \in \mathbb{R} \\ \theta_{n+1}(x) = \theta_n(x) - \frac{1}{n^\gamma} \left(\mathbf{1}_{Y_{n+1} \leq \theta_n(x)} - \alpha \right) \mathbf{1}_{X_{n+1} \in kNN_{n+1}(x)} \end{cases}$$

où $k_n = \lfloor n^\beta \rfloor$.

Un algorithme stochastique plongé dans la méthode des k -plus proches voisins

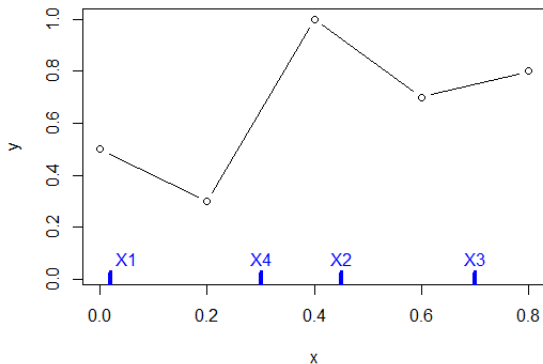
L'algorithme

$$\begin{cases} \theta_0(x) \in \mathbb{R} \\ \theta_{n+1}(x) = \theta_n(x) - \frac{1}{n^\gamma} \left(\mathbf{1}_{Y_{n+1} \leq \theta_n(x)} - \alpha \right) \mathbf{1}_{X_{n+1} \in kNN_{n+1}(x)} \end{cases}$$

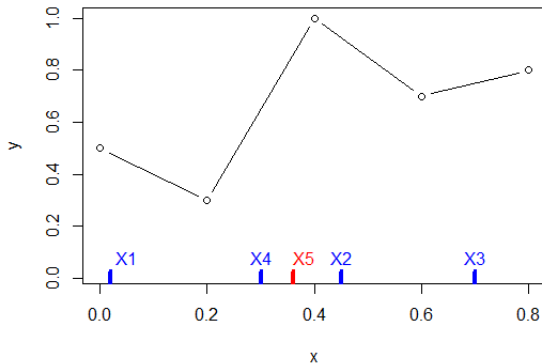
où $k_n = \lfloor n^\beta \rfloor$.

- Pour quels paramètres (γ, β) l'algorithme **se comporte-il bien** ? $\Rightarrow$ besoin d'un **compromis**.
- Peut-on montrer des résultats **non-asymptotiques** ?
- Y-a-t-il des **paramètres optimaux** ?

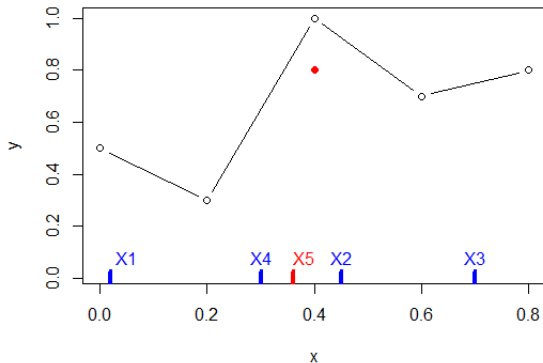
Exemple



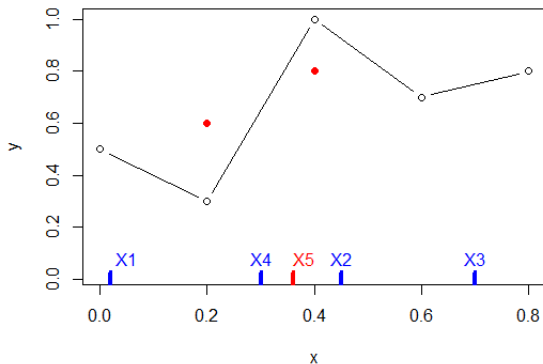
Exemple



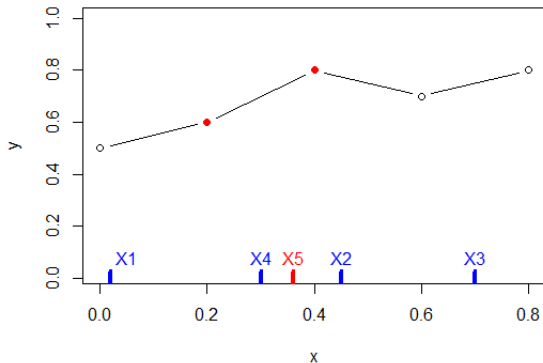
Exemple



Exemple



Exemple



- 1 Estimation empirique
 - Le superquantile
 - Une nouvelle mesure de risque : le superquantile de Bregman
- 2 Estimation séquentielle par algorithme stochastique et méthode des k plus proches voisins
 - Construction de l'algorithme
 - **Résultats**
 - Simulations numériques
- 3 Estimation séquentielle bayésienne et planification d'expériences
 - Etat de l'art
 - Deux nouvelles stratégies
 - Résultats numériques

Hypothèse de continuité

Notations :

- $\mathcal{B}_x$ l'ensemble des boules de $\mathbb{R}^d$ centrées en x . $B \in \mathcal{B}_x$ a pour rayon r_B .
- Pour $B \in \mathcal{B}_x$, F_Y^B est la fonction de répartition de $\mathcal{L}(g(X, \epsilon) | X \in B)$.
- F_{Y^x} est la fonction de répartition de la loi de $\mathcal{L}(g(x, \epsilon))$.

Hypothèse A1 : Pour tout $x \in \text{Supp}(X)$, il existe une constante $M(x)$ telle que

$$\forall B \in \mathcal{B}_x, \forall t \in \mathbb{R}, |F_{Y^B}(t) - F_{Y^x}(t)| \leq M(x)r_B.$$

Hypothèses techniques

Hypothèse A2 : La loi d'entrée X a une densité minorée par une constante $C_{inputs} > 0$.

$\Rightarrow$ Utile pour gérer $\mathbb{E}(\|X - x\|_{(k_n, n)})$ où $\mathbb{P}(X \in kNN_n(x))$.

Hypothèses techniques

Hypothèse A2 : La loi d'entrée X a une densité minorée par une constante $C_{inputs} > 0$.

$\Rightarrow$ Utile pour gérer $\mathbb{E}(\|X - x\|_{(k_n, n)})$ où $\mathbb{P}(X \in kNN_n(x))$.

Hypothèse A3 : La fonction code g est à valeur dans le compact $[L_Y, U_Y]$.

$\Rightarrow \forall x, \theta_n(x)$ est presque-sûrement borné.

Hypothèses techniques

Hypothèse A2 : La loi d'entrée X a une densité minorée par une constante $C_{inputs} > 0$.

$\Rightarrow$ Utile pour gérer $\mathbb{E}(\|X - x\|_{(k_n, n)})$ où $\mathbb{P}(X \in kNN_n(x))$.

Hypothèse A3 : La fonction code g est à valeur dans le compact $[L_Y, U_Y]$.

$\Rightarrow \forall x, \theta_n(x)$ est presque-sûrement borné.

Hypothèse A4 : Pour tout $x \in \text{Supp}(X)$, la loi de $g(x, \epsilon)$ est à densité minorée par une constante $C_g(x) > 0$.

$\Rightarrow$ Il existe une constante $C_2(x, \alpha)$ telle que :

$$\forall \theta_n(x), [F_{Y^x}(\theta_n(x)) - F_{Y^x}(\theta^*(x))] [\theta_n(x) - \theta^*(x)] \geq C_2(x, \alpha) [\theta_n(x) - \theta^*(x)]^2.$$

Vitesse de convergence de l'erreur quadratique

Proposition (L.-R. et al. - 2016)

Sous les hypothèses **A1**, **A2**, **A3** et **A4**, la MSE $a_n(x)$ satisfait :
 $\forall(\gamma, \beta, \epsilon)$ tels que $0 < \gamma \leq \beta < 1$ et $1 > \epsilon > 1 - \beta$, $\forall n \geq N_0$,

$$a_n(x) \leq \exp \left(-2C_2(x, \alpha) \sum_{j=N_0+1}^n j^{-\epsilon-\gamma} \right) C_1$$

$$+ \sum_{k=N_0+1}^n \exp \left(-2C_2(x, \alpha) \sum_{j=k+1}^n j^{-\epsilon-\gamma} \right) d_k + C_1 \exp \left(-\frac{3n^{1-\epsilon}}{8} \right)$$

$$\text{où } d_n = C_1 \exp \left(-\frac{3n^{1-\epsilon}}{8} \right) + 2\sqrt{C_1} M(x) C_3(d) \gamma_n \left(\frac{k_n}{n} \right)^{\frac{1}{d}+1} + \gamma_n^2 \frac{k_n}{n}.$$

Compromis entre les deux erreurs

- L'erreur **d'apprentissage** donne le terme

$$\exp \left(-2C_2(x, \alpha)(x) \sum_{k=N_0+1}^n \frac{1}{k^{\epsilon+\gamma}} \right).$$

Ce terme décroît à 0 si, et seulement si $\beta \geq \gamma$. Ainsi, β **ne doit pas être choisi trop petit**.

- L'erreur **de biais** donne le terme

$$\gamma_n (k_n/n)^{1/d+1} = n^{(1-\beta)(1+1/d)+\gamma}.$$

Il faut donc que $\beta < 1$. Donc β **ne doit pas être choisi trop grand**.

Pourquoi faut-il que $\gamma \leq \beta$?

Algorithme classique :

- Une mise à jour à chaque pas
- $\sum_n \gamma_n = +\infty \Rightarrow \gamma \leq 1$.

Pourquoi faut-il que $\gamma \leq \beta$?

Algorithme classique :

- Une mise à jour à chaque pas
- $\sum_n \gamma_n = +\infty \Rightarrow \gamma \leq 1.$

Notre algorithme :

- Une mise à jour tous les $\frac{k_n}{n} \sim n^{\beta-1}$ pas.
- Au pas n : $N = \sum_{k \leq n} k^{\beta-1} \sim n^\beta$ mises à jour.
- Au temps $t = n^{\frac{1}{\beta}}$: $N = n$ mises à jour.
- **Taux d'apprentissage effectif** : $\gamma_{k_n} = \frac{1}{\left(n^{\frac{1}{\beta}}\right)^\gamma}.$
- $\sum_n \gamma_{k_n} = +\infty \Rightarrow \gamma \leq \beta.$

Corollaire : paramètres optimaux (L.-R. et al. - 2016)

Sous les mêmes hypothèses, l'erreur quadratique décroît le plus rapidement lorsque les paramètres sont $\gamma = \frac{1}{1+d}$ et $\beta = \gamma + \eta_\beta$ où $\eta_\beta > 0$ est aussi petit que possible. De plus, avec ces paramètres, il existe **une constante** $C_9(x, \alpha, d)$ telle que $\forall n \geq N_4(x, \alpha, d)$,

$$a_n(x) \leq \frac{C_9(x, \alpha, d)}{n^{\frac{1}{1+d} - \eta}}$$

où $\eta = \frac{\eta_\epsilon}{2} + \eta_\beta$ et $\eta_\epsilon = 1 - \beta - \epsilon$.

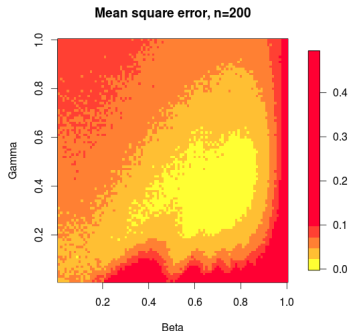
- 1 Estimation empirique
 - Le superquantile
 - Une nouvelle mesure de risque : le superquantile de Bregman
- 2 Estimation séquentielle par algorithme stochastique et méthode des k plus proches voisins
 - Construction de l'algorithme
 - Résultats
 - Simulations numériques
- 3 Estimation séquentielle bayésienne et planification d'expériences
 - Etat de l'art
 - Deux nouvelles stratégies
 - Résultats numériques

Dimension 1

Modèle :

$$g(X, \epsilon) = X^2 + \epsilon$$

avec $X \sim \mathcal{U}([-1, 1])$, $\epsilon \sim \mathcal{U}([-0.5, 0.5])$, $x = 0$ et $\alpha = 0.95$.

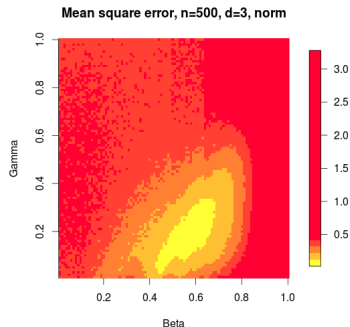
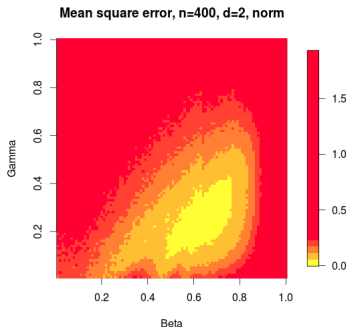


Dimensions 2 et 3

Modèle :

$$g(X, \epsilon) = ||X||^2 + \epsilon$$

avec $X \sim \mathcal{U}([-1, 1]^d)$, $\epsilon \sim \mathcal{U}([-0.5, 0.5])$, $x = 0_{\mathbb{R}^d}$ et $\alpha = 0,95$.



Bilan

Conclusion :

- Nous avons introduit un algorithme pour **estimer le quantile de la sortie d'un code stochastique coûteux**.
- Nous avons donné les **meilleurs paramètres** pour faire tourner l'algorithme.
- Les simulations numériques confirment que notre algorithme est performant pour résoudre le problème posé.

Perspectives :

- Que se passe-t-il si on relâche les hypothèses de **support compact** ?
- Peut-on écrire des **bornes inférieures** ?
- Appliquer cet algorithme à de vrais jeux de données (fournis par EDF).

- 1 Estimation empirique
 - Le superquantile
 - Une nouvelle mesure de risque : le superquantile de Bregman
- 2 Estimation séquentielle par algorithme stochastique et méthode des k plus proches voisins
 - Construction de l'algorithme
 - Résultats
 - Simulations numériques
- 3 Estimation séquentielle bayésienne et planification d'expériences
 - Etat de l'art
 - Deux nouvelles stratégies
 - Résultats numériques

Stratégie séquentielle de planification d'expériences

CADRE : estimation de **quantiles** en **code numérique très coûteux**.

Stratégie de planification d'expériences

- Pour un budget initial N_0 , on construit un **échantillon d'initialisation** $(\mathbf{x}_0^i, g(\mathbf{x}_0^i))_{i=1\dots N_0} = \mathcal{A}_0$, et on calcule **l'estimateur initial du quantile** q_{N_0} .
- A chaque étape $n + 1 \geq N_0 + 1$ et jusqu'à ce que le budget d'évaluations autorisées N soit atteint : **connaissant les précédentes évaluations $\mathcal{A}_n$ et l'estimateur q_n , on choisit le futur point à évaluer $\mathbf{x}_{n+1}^*$ grâce à un certain critère.** On évalue $g(\mathbf{x}_{n+1}^*) := g_{n+1}$ et on met à jour $\mathcal{A}_{n+1}$ et q_{n+1} .
- q_N est l'estimateur du quantile à retourner.

Stratégie séquentielle de planification d'expériences

CADRE : estimation de **quantiles** en **code numérique très coûteux**.

Stratégie de planification d'expériences

- Pour un budget initial N_0 , on construit un **échantillon d'initialisation** $(\mathbf{x}_0^i, g(\mathbf{x}_0^i))_{i=1\dots N_0} = \mathcal{A}_0$, et on calcule **l'estimateur initial du quantile** q_{N_0} .
- A chaque étape $n + 1 \geq N_0 + 1$ et jusqu'à ce que le budget d'évaluations autorisées N soit atteint : **connaissant les précédentes évaluations $\mathcal{A}_n$ et l'estimateur q_n , on choisit le futur point à évaluer $\mathbf{x}_{n+1}^*$ grâce à un certain critère.** On évalue $g(\mathbf{x}_{n+1}^*) := g_{n+1}$ et on met à jour $\mathcal{A}_{n+1}$ et q_{n+1} .
- q_N est l'estimateur du quantile à retourner.

BESOIN : **un estimateur séquentiel du quantile** et un **critère de sélection du nouveau point à évaluer**.

- 1 Estimation empirique
 - Le superquantile
 - Une nouvelle mesure de risque : le superquantile de Bregman
- 2 Estimation séquentielle par algorithme stochastique et méthode des k plus proches voisins
 - Construction de l'algorithme
 - Résultats
 - Simulations numériques
- 3 Estimation séquentielle bayésienne et planification d'expériences
 - Etat de l'art
 - Deux nouvelles stratégies
 - Résultats numériques

Hypothèse gaussienne

Modèle : g est la réalisation d'un processus gaussien G centré et de fonction de covariance $c(.,.)$ fixée.

Propriété : krigeage universel

Pour $\mathcal{A}_n = \{(x_1, g_1 := g(x_1)), \dots, (x_n, g(x_n) := g_n)\}$, on sait que

$$\mathcal{L}(G|\mathcal{A}_n) = GP(m_n(.), k_n(.,.)),$$

où $\forall \mathbf{x} \in \mathbb{X}$,

$$m_n(\mathbf{x}) = \mathbb{E}(G(\mathbf{x})|\mathcal{A}_n) = c_n(\mathbf{x})^T C_n^{-1} \mathbf{g}_n,$$

$$k_n(\mathbf{x}, \mathbf{x}') = \text{Cov}(G(\mathbf{x}), G(\mathbf{x}')|\mathcal{A}_n) = c(\mathbf{x}, \mathbf{x}') - c_n(\mathbf{x})^T C_n^{-1} c_n(\mathbf{x}'),$$

avec $c_n(\mathbf{x}) = [c(\mathbf{x}_1, \mathbf{x}), \dots, c(\mathbf{x}_n, \mathbf{x})]^T$, $C_n = [c(\mathbf{x}_i, \mathbf{x}_j)]_{1 \leq i, j \leq n}$ et $\mathbf{g}_n = [g_1, \dots, g_n]$.

Deux estimateurs du quantile

- **Conditionnellement à $\mathcal{A}_n$, la meilleure approximation de $G(\mathbf{x})$ est $m_n(\mathbf{x})$** , un estimateur intuitif est donc le quantile de la moyenne du processus gaussien.

$$q_n := q(m_n(X)) = q(\mathbb{E}[G(X)|\mathcal{A}_n]). \quad (4)$$

- L'estimateur **minimisant l'erreur en moyenne quadratique $\mathbb{E}((q - q_n)^2)$ parmi tous les estimateurs $\mathcal{A}_n$ -mesurables** est

$$q_n := \mathbb{E}(q(G(X))|\mathcal{A}_n). \quad (5)$$

Critère de sélection et méthode SUR

- **Stepwise Uncertainty Reduction** : Geman et al. (1996).
- On recherche des critères de la forme

$$\mathbf{x}_{n+1}^* := \operatorname{argmin}_{\mathbf{x}_{n+1} \in \mathbb{X}} J_n(\mathbf{x}_{n+1}),$$

- Difficulté : **évaluer l'impact potentiel du point candidat $\mathbf{x}_{n+1}$ sans avoir accès à son évaluation g_{n+1} .**

Application au quantile : Arnaud et al. (2010), Jala et al. (2012)

⇒ Minimisation de la variance conditionnelle de l'estimateur (5).

⇒ Méthode performante en petite dimension mais **trop coûteuse**.

- 1 Estimation empirique
 - Le superquantile
 - Une nouvelle mesure de risque : le superquantile de Bregman
- 2 Estimation séquentielle par algorithme stochastique et méthode des k plus proches voisins
 - Construction de l'algorithme
 - Résultats
 - Simulations numériques
- 3 Estimation séquentielle bayésienne et planification d'expériences
 - Etat de l'art
 - **Deux nouvelles stratégies**
 - Résultats numériques

Une expression *explicite* de l'estimateur (4)

- L'estimateur (1) est $q_n := q(m_n(X))$.
- Grâce à un nouvel échantillon $\mathbf{X}_{\text{MC}} = (\mathbf{x}_{\text{MC}}^1, \dots, \mathbf{x}_{\text{MC}}^I)$ de X , on a accès à une **version empirique**

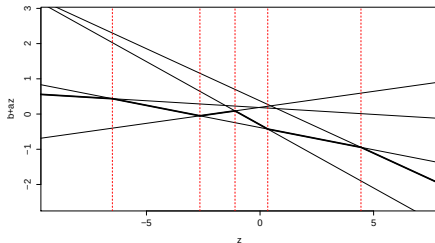
$$q_n = m_n(\mathbf{X}_{\text{MC}})_{(\lfloor I\alpha \rfloor + 1)} := m_n(\mathbf{x}_n^q).$$

- **BUT** : **exprimer** $q_{n+1} = m_{n+1}(\mathbf{x}_{n+1}^q)$ **en fonction du modèle à l'étape n et du nouveau couple de données $(\mathbf{x}_{n+1}, g_{n+1})$.**

Formule de mise à jour à un pas (Chevalier et al. -2014)

$$m_{n+1}(\mathbf{X}_{\text{MC}}) = m_n(\mathbf{X}_{\text{MC}}) + \frac{k_n(\mathbf{X}_{\text{MC}}, \mathbf{x}_{n+1})}{s_n(\mathbf{x}_{n+1})^2} (g_{n+1} - m_n(\mathbf{x}_{n+1})).$$

Une expression *explicite* de l'estimateur (1)



Proposition (L.-R. et al. - 2016)

Il existe une suite d'intervalles (B_i) connus telle que

$$q_{n+1}(\mathbf{x}_{n+1}, g_{n+1}) = \sum_{i=0}^L m_{n+1}(\mathbf{x}_{n+1}^q(B_i)) \mathbf{1}_{g_{n+1} \in B_i}$$

Le critère *probabilité*

- Par définition du quantile, on a

$$\mathbb{P}(G(X) \geq q(G(X))) = 1 - \alpha.$$

- On propose donc le critère suivant :

$$J_n^{\text{prob}} = \left| \int_{\mathbb{X}} \mathbb{P}(G(\mathbf{x}) \geq q_n | \mathcal{A}_n) d\mathbf{x} - (1 - \alpha) \right| = |\Gamma_n - (1 - \alpha)|,$$

avec $\Gamma_n = \int_{\mathbb{X}} \mathbb{P}(G(\mathbf{x}) \geq q_n | \mathcal{A}_n) d\mathbf{x}$.

- PROBLEME** : $J_n^{\text{prob}}(x_{n+1})$ dépend de g_{n+1} .

- **SOLUTION** : Sous l'hypothèse gaussienne et à l'étape n , g_{n+1} est la réalisation d'une variable aléatoire $G_{n+1} \sim \mathcal{N}(m_n(x_{n+1}), s_n^2(x_{n+1})) \Rightarrow$ on moyennise.
- Le critère devient alors

$$J_n^{\text{prob}}(\mathbf{x}_{n+1}) = |\mathbb{E}_{G_{n+1}}(\Gamma_{n+1}(x_{n+1})) - (1 - \alpha)|$$

avec

$$\Gamma_{n+1}(x_{n+1}) = \int_{\mathbb{X}} \mathbb{P}(G(\mathbf{x}) \geq q_{n+1} | A_{n+1}) d\mathbf{x},$$

et $A_{n+1} = \mathcal{A}_{n+1} \cup (\mathbf{x}_{n+1}, G_{n+1})$.

Proposition (L.-R. et al. - 2016)

$$J_n^{\text{prob}}(\mathbf{x}_{n+1}) = \left| \int_{\mathbb{X}} \sum_{i=1}^{L-1} \left[\Phi_{r_i^n} \left(e_n^i(\mathbf{x}_{n+1}; \mathbf{x}), f_n^i(\mathbf{x}_{n+1}, l_{i+1}) \right) \right. \right. \\
- \Phi_{r_i^n(\mathbf{x}_{n+1}, \mathbf{x})} \left(e_n^i(\mathbf{x}_{n+1}; \mathbf{x}), f_n(\mathbf{x}_{n+1}, l_i) \right) \\
+ \Phi_{r_i^n} \left((e_n^i(\mathbf{x}_{n+1}; \mathbf{x}), f_n^i(\mathbf{x}_{n+1}, l_1)) \right) \\
\left. \left. + \Phi_{-r_i^n} \left(e_n^i(\mathbf{x}_{n+1}; \mathbf{x}), -f_n^i((\mathbf{x}_{n+1}, l_L)) \right) \right] d\mathbf{x} - (1 - \alpha) \right|$$

Le critère *variance*

- q_n doit converger vers le quantile $\Rightarrow$ on veut un estimateur **de plus en plus *stable***.
- **La variance de $q_{n+1} | \mathcal{A}_n \cup (\mathbf{x}_{n+1}, G_{n+1})$ est un bon indicateur de *stabilité*.**
- Choix du point qui maximise cette variance pour obtenir le point dont l'évaluation a un large impact sur la valeur de l'estimateur $\Rightarrow$ **Réduction de l'instabilité sur l'estimateur.**

$$J_n^{\text{Var}}(\mathbf{x}_{n+1}) = \text{Var}_{G_{n+1}}(q_{n+1} | \mathcal{A}_{n+1})$$

avec $\mathcal{A}_{n+1} = \mathcal{A}_n \cup (\mathbf{x}_{n+1}, G_{n+1})$.

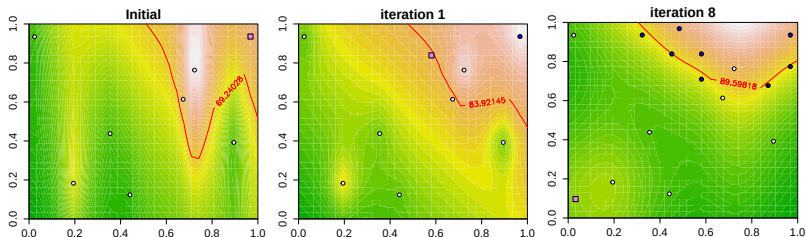
Proposition (L.-R. et al. - 2016)

$$\begin{aligned}
 J_n^{\text{Var}}(\mathbf{x}_{n+1}) = & \sum_{i=1}^L \left[k_n(\mathbf{x}_{n+1}^q(B_i), \mathbf{x}_{n+1}) \right]^2 V(s_n(\mathbf{x}_{n+1}), l_{i+1}, l_i) P_i \\
 & + \sum_{i=1}^L \left[m_n(\mathbf{x}_{n+1}^q(B_i) - k_n(\mathbf{x}_{n+1}^q(B_i), \mathbf{x}_{n+1}) E(s_n(\mathbf{x}_{n+1}), l_{i+1}, l_i) \right]^2 (1 - P_i) P_i \\
 & - 2 \sum_{i=2}^L \sum_{j=1}^{i-1} \left[m_n(\mathbf{x}_{n+1}^q(B_i) - k_n(\mathbf{x}_{n+1}^q(B_i), \mathbf{x}_{n+1}) E(s_n(\mathbf{x}_{n+1}), l_{i+1}, l_i) \right] P_i \\
 & \left[m_n(\mathbf{x}_{n+1}^q(B_j) - k_n(\mathbf{x}_{n+1}^q(B_j), \mathbf{x}_{n+1}) E(s_n(\mathbf{x}_{n+1}), l_{j+1}, l_j) \right] P_j,
 \end{aligned}$$

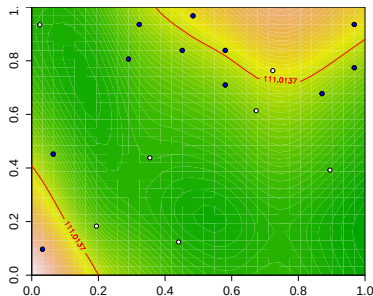
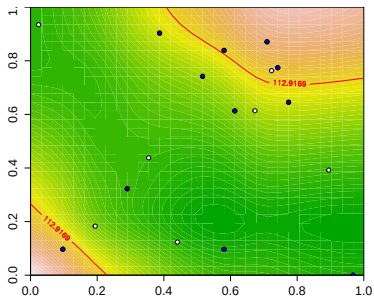
- 1 Estimation empirique
 - Le superquantile
 - Une nouvelle mesure de risque : le superquantile de Bregman
- 2 Estimation séquentielle par algorithme stochastique et méthode des k plus proches voisins
 - Construction de l'algorithme
 - Résultats
 - Simulations numériques
- 3 Estimation séquentielle bayésienne et planification d'expériences
 - Etat de l'art
 - Deux nouvelles stratégies
 - Résultats numériques

Critère *variance* en dimension 2

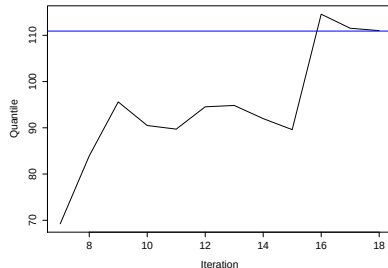
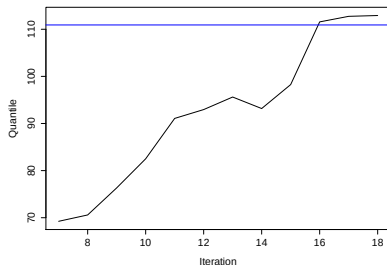
Modèle : $Y = g(X)$ avec g la fonction de Branin, $X \sim \mathcal{U}([0, 1]^2)$,
fonction de covariance Matérn $3/2$ et $\alpha = 0.85$.



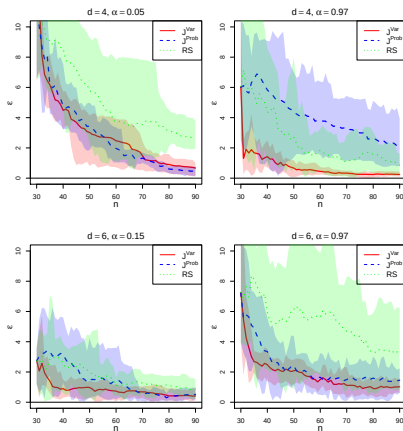
Comparaison des deux critères en dimension 2



Convergence des deux estimateurs en dimension 2



En dimension supérieure



Bilan

Conclusion :

- Nous avons donné **deux nouvelles stratégies** basées sur des critères ayant des **formules analytiques** pour estimer le quantile de la sortie d'un **code numérique très coûteux**.
- Les simulations numériques montrent que nos stratégies sont efficaces même lorsque **la dimension des entrées augmente**.

Perspectives :

- Comparer nos stratégies avec les précédentes en petite dimension.
- Trouver des manières de **minimiser le critère moins coûteuses en temps de calcul**.
- Démontrer des **garanties théoriques** pour ces stratégies.

Merci pour votre attention.

COMPLEMENTS

Exemples et contre-exemples de cohérence de SQB

- **Fonction Bregman géométrique** : $\gamma'(x) = x \mapsto \ln(x)$.
 - i) et ii) satisfaits.
 - γ' est concave et sous-additive seulement sur $[1, +\infty[$. Ainsi, iii) est satisfait pour les couples (X, X') tels que,

$$\min \left(q_{\alpha}^X(\alpha), q_{\alpha}^{X'}(\alpha), q_{\alpha}^{X+X'}(\alpha) \right) > 1$$

- **Le sous-additivité et l'homogénéité sont fausses dans le cas général.** Pour $\gamma(x) = \exp(x)$ et $X \sim \mathcal{U}([0, 1])$:

$$\mathcal{R}(2X) - 2\mathcal{R}(X) = 0.000107 > 0,$$

et

$$\frac{\mathcal{R}(4X)}{4\mathcal{R}(X)} = 1,000321 > 1$$

Exemples de moyennes de Bregman

- **Euclidien.** $\gamma(x) = x^2$ sur $\mathbb{R}$, on obtient, pour $x, x' \in \mathbb{R}$,

$$d_\gamma(x, x') = (x - x')^2$$

et b est l'**espérance classique**.

- **Géométrique.** $\gamma(x) = x \ln(x) - x + 1$ sur $\mathbb{R}_+^*$ on obtient, pour $x, x' \in \mathbb{R}_+^*$,

$$d_\gamma(x, x') = x \ln \frac{x}{x'} + x' - x$$

et b est la **moyenne géométrique**.

- **Harmonique.** $\gamma(x) = -\ln(x) + x - 1$ sur $\mathbb{R}_+^*$ on obtient, pour $x, x' \in \mathbb{R}_+^*$,

$$d_\gamma(x, x') = -\ln \frac{x}{x'} + \frac{x}{x'} - 1$$

et b est la **moyenne harmonique**.

Exemples de SQB

- Géométrique. $\gamma(x) = x \ln(x) - x + 1$ sur $\mathbb{R}_+^*$ on obtient,

$$\exp \left[\mathbb{E} \left(\frac{\ln(X) \mathbf{1}_{X \geq F_X^{-1}(\alpha)}}{1 - \alpha} \right) \right]$$

- Harmonique. $\gamma(x) = -\ln(x) + x - 1$ sur $\mathbb{R}_+^*$ on obtient,

$$\frac{1}{1 - \left[\mathbb{E} \left(\frac{\left(-\frac{1}{X} + 1\right) \mathbf{1}_{X \geq F_X^{-1}(\alpha)}}{1 - \alpha} \right) \right]}$$

Estimation séquentielle par algorithme stochastique (Robbins et Monro - 1951)

- Pour approcher le zéro de $h : \mathbb{R}^d \rightarrow \mathbb{R}$, on peut appliquer un algorithme du gradient.
- Lorsque la fonction h dépend d'une loi inconnue Z , on construit un échantillon $(Z_1, \dots, Z_n)$ et on applique un algorithme stochastique :

$$\begin{cases} \theta_0 \in \mathbb{R}^d \\ \theta_{n+1} = \theta_n - \gamma_{n+1} H(\theta_n, Z_{n+1}) \end{cases} \quad (6)$$

où $\mathcal{F}_n := \sigma(Z_1, \dots, Z_n)$, $\mathbb{E}(H(\theta_n, Z_{n+1}) | \mathcal{F}_n) = h(\theta_n)$,

et (γ_n) est une suite déterministe de pas bien choisie.

Le cas du quantile dans un code numérique

Dans notre cas,

- a) On évalue un échantillon $(Y_1, \dots, Y_n)$ en fournissant $(X_1, \dots, X_n)$ au code.
- b) Puisque q_α est la zéro de la fonction $F_Y - \alpha$, on applique l'algorithme

$$\begin{cases} q_0 \in \mathbb{R} \\ q_{n+1} = q_n - \gamma_{n+1} (\mathbf{1}_{Y_{n+1} \leq q_n} - \alpha) \end{cases} \quad (7)$$

où $\gamma_n = n^{-\gamma}$ avec $\gamma \in]0, 1]$.

- c) Question : comment régler le paramètre γ ?

Comportement asymptotique

Théorème de Robbins-Monro (1951)- Consistance

Si $f_Y(q) \neq 0$ et si la suite de pas déterministes $(\gamma_n)_n$ vérifie

$$\sum_{n \geq 0} \gamma_n = \infty \text{ et } \sum_{n \geq 0} \gamma_n^2 < \infty$$

alors l'algorithme (7) converge presque sûrement vers q .

- Conclusion : il faut donc choisir $\gamma \in]1/2, 1]$.

Théorème de Robbins-Monro (1951)

Sous des hypothèses de régularité sur la loi Y , lorsque $\gamma \in]\frac{1}{2}, 1]$, l'algorithme produit un estimateur normalement asymptotique à vitesse $\sqrt{n^\gamma}$

- Conclusion : il faut donc choisir γ le plus proche possible de 1. ↻ 🔍 🔗

Comportement asymptotique

Théorème de Robbins-Monro (1951)

Si f_Y est $\mathcal{C}^1$ et si $f(q) > 0$, alors pour $\gamma_n = n^{-1}$, l'algorithme (7) vérifie

1) Si $f(q) > \frac{1}{2}$, alors

$$\sqrt{n}(q_n - q) \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, \frac{\alpha(1-\alpha)}{2f(q) - 1}\right).$$

2) Si $f(q) = \frac{1}{2}$,

$$\sqrt{\frac{n}{\ln(n)}}(q_n - q) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \alpha(1-\alpha)).$$

3) Si $f(q) < \frac{1}{2}$,

$$n^{f(q)}(q_n - q) \xrightarrow{p.s.} Z$$

Hypothèse H1 : La variable aléatoire Y est à densité minorée par une constante $D > 0$. Donc :

$$(q_n - q)(F(q_n) - F(q)) \geq D(q_n - q)^2. \quad (8)$$

Lien entre D et $f(q)$

Comme

$$h(q_n) = |F(q_n) - F(q)| \sim f(q)|q_n - q|,$$

D "joue le rôle de $f(q)$ ".

Proposition

Sous l'hypothèse (8), le risque quadratique de l'algorithme (7) vérifie les inégalités non-asymptotiques suivantes.

Dans le cas où $\gamma_n = \frac{1}{n}$, pour tout $n \geq 1$, on a

$$a_n \leq \begin{cases} n^{-2D} a_0 + \frac{4}{(2D-1)n} + \frac{1}{n^2} & \text{si } D > \frac{1}{2} \\ (a_0 + 4C_1)n^{-2D} + \frac{1}{n^2} & \text{si } D < \frac{1}{2} \\ \frac{a_0}{n} + \frac{4 \ln(n)}{n} + \frac{1}{n^2} & \text{si } D = \frac{1}{2}, \end{cases}$$

Proposition

Dans le cas où $\gamma_n = n^{-\gamma}$ avec $\gamma \in]\frac{1}{2}, 1[$, pour tout $n \geq \max(n_1, 3)$ on a

$$a_n \leq \frac{2^{2\gamma-1}}{Dn^\gamma} + \exp\left(-Dn^{-\gamma+1}\right) C_2 \frac{(n+1)^{-2\gamma+1}}{-2\gamma+1} \\ + \exp\left(-2D \frac{n^{-\gamma+1}}{-\gamma+1}\right) a_0,$$

Les résultats sont cohérents avec le TCL

1) Lorsque $f(q) > \frac{1}{2}$, asymptotiquement en loi, nous avons

$$q_n \sim q + \frac{\sigma}{\sqrt{n}} \mathcal{N}(0, 1).$$

Notre résultat nous donne parallèlement que si $D > \frac{1}{2}$, alors

$$\mathbb{E} \left[(q_n - q)^2 \right] \sim C_{ste} \frac{1}{n}.$$

Les résultats sont cohérents avec le TCL

2) Lorsque $f(q) = \frac{1}{2}$, asymptotiquement et lois, nous avons

$$q_n \sim q + \sigma' \sqrt{\frac{\log(n)}{n}} \mathcal{N}(0, 1).$$

Notre résultat nous donne parallèlement que si $D = \frac{1}{2}$, alors

$$\mathbb{E} \left[(q_n - q)^2 \right] \sim C_{ste} \frac{\log(n)}{n}.$$

Les résultats sont cohérents avec le TCL

3) Lorsque $f(q) < \frac{1}{2}$, asymptotiquement en loi, nous avons

$$q_n \sim q + \frac{\sigma}{n^{f(q)}} Z.$$

Notre résultat nous donne parallèlement que si $D < \frac{1}{2}$, alors

$$\mathbb{E} \left[(q_n - q)^2 \right] \sim C_{ste} \frac{1}{n^{2D}}.$$

Difficultés pour généraliser le résultat

Comment remplacer l'hypothèse H1 ?

- Lorsque q_n est "loin" de sa cible ($q = 0$), c'est à dire si $|q_n| \geq \epsilon$,

$$h(q_n) = \int_0^{q_n} f(x)dx \geq \int_0^{\epsilon} f(x)dx \geq \frac{f(0)}{2}\epsilon.$$

On ne sait pas majorer $q_n h(q_n)$ par une quantité de l'ordre de q_n^2 .

- Lorsque q_n est proche de sa cible, c'est à dire lorsque $|q_n| \leq \epsilon$,

$$f(q_n)q_n \geq q_n^2 f(0) - \|f'\|_{\infty} \frac{q_n^3}{2} \geq q_n^2 f(0) \left(1 - \frac{\epsilon \|f'\|_{\infty}}{2f(0)}\right).$$

Idée de raisonnement

On remarque d'abord que

$$a_n \leq \mathbb{E} \left[q_n^2 \mathbf{1}_{q_n \leq \epsilon_n} \right] + \mathbb{E} \left[q_n^2 \mathbf{1}_{q_n \geq \epsilon_n} \right] := b_n + \bar{b}_n.$$

- On peut montrer que

$$b_{n+1} \leq b_n(1 - c_{n+1}) + \gamma_{n+1}^2 + (q_0 + \alpha \Gamma_n)^2 \mathbb{P}(q_n \geq \epsilon_n)$$

avec $c_n = 2f(0)(1 - C\epsilon_n)$.

- On majore

$$\bar{b}_n \leq (q_0 + \alpha \Gamma_n)^2 \mathbb{P}(q_n \geq \epsilon_n).$$

$\Rightarrow$ Besoin d'un **compromis sur ϵ_n** : DIFFICILE !

Convergence p.s.

Proposition : Convergence presque-sûre (L.-R. et al. - 2016)

Soit x une entrée fixée. Sous les hypothèses **A1** et **A2**, l'algorithme en x est presque-sûrement convergent dès que $\frac{1}{2} < \gamma < \beta < 1$.

Convergence p.s.

Proposition : Convergence presque-sûre (L.-R. et al. - 2016)

Soit x une entrée fixée. Sous les hypothèses **A1** et **A2**, l'algorithme en x est presque-sûrement convergent dès que $\frac{1}{2} < \gamma < \beta < 1$.

Commentaires sur les paramètres :

- $\frac{1}{2} < \gamma \leq 1 \Rightarrow$ hypothèses classiques pour les algorithmes de Robbins-Monro.

Convergence p.s.

Proposition : Convergence presque-sûre (L.-R. et al. - 2016)

Soit x une entrée fixée. Sous les hypothèses **A1** et **A2**, l'algorithme en x est presque-sûrement convergent dès que $\frac{1}{2} < \gamma < \beta < 1$.

Commentaires sur les paramètres :

- $\frac{1}{2} < \gamma \leq 1 \Rightarrow$ hypothèses classiques pour les algorithmes de Robbins-Monro.
- $0 < \beta \Rightarrow$ le nombre de voisins tend vers $+\infty$ et donc,
 $\|X - x\|_{(k_n, n)} \rightarrow 0$.

Convergence p.s.

Proposition : Convergence presque-sûre (L.-R. et al. - 2016)

Soit x une entrée fixée. Sous les hypothèses **A1** et **A2**, l'algorithme en x est presque-sûrement convergent dès que $\frac{1}{2} < \gamma < \beta < 1$.

Commentaires sur les paramètres :

- $\frac{1}{2} < \gamma \leq 1 \Rightarrow$ hypothèses classiques pour les algorithmes de Robbins-Monro.
- $0 < \beta \Rightarrow$ le nombre de voisins tend vers $+\infty$ et donc,
 $\|X - x\|_{(k_n, n)} \rightarrow 0$.
- $\beta < 1 \Rightarrow$ hypothèse technique.

Convergence p.s.

Proposition : Convergence presque-sûre (L.-R. et al. - 2016)

Soit x une entrée fixée. Sous les hypothèses **A1** et **A2**, l'algorithme en x est presque-sûrement convergent dès que $\frac{1}{2} < \gamma < \beta < 1$.

Commentaires sur les paramètres :

- $\frac{1}{2} < \gamma \leq 1 \Rightarrow$ hypothèses classiques pour les algorithmes de Robbins-Monro.
- $0 < \beta \Rightarrow$ le nombre de voisins tend vers $+\infty$ et donc,
 $\|X - x\|_{(k_n, n)} \rightarrow 0$.
- $\beta < 1 \Rightarrow$ hypothèse technique.
- $\gamma < \beta \Rightarrow$ *taux d'apprentissage effectif*.

Esquisse de preuve

- 1) On décompose H en un terme de martingale et un reste, en posant :

$$h_n(q_n) = \mathbb{E}(H(q_n, X_{n+1}, Y_{n+1}) | \mathcal{F}_n) \text{ et } \xi_{n+1} = H(q_n, X_{n+1}, Y_{n+1}) - h_n(q_n).$$

On a

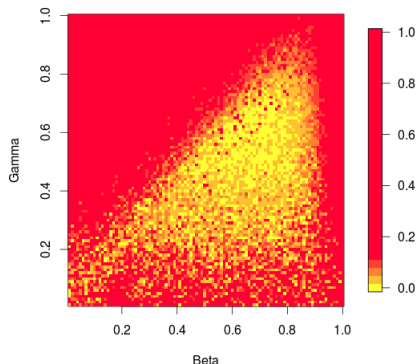
$$T_n = q_n(x) + \sum_{j=1}^n \gamma_j h_{j-1}(q_{j-1}(x))$$

martingale bornée dans L^2 donc qui converge presque-sûrement.

- 2) On montre la convergence presque-sûre de $(q_n(x))_n$.
 - a) $(q_n(x))$ ne diverge pas vers $+\infty$ ou $-\infty$.
 - b) $(q_n(x))$ converge p.s vers une limite finie.
- 3) La limite est $q(x)$ le quantile conditionnel que nous voulons estimer.

Illustration numérique de la convergence p. s.

relative error in function of Beta and Gamma, $n=5000$



Esquisse de preuve

- L'idée de la preuve est d'établir une inégalité du genre

$$a_{n+1}(x) \leq a_n(x)(1 - c_{n+1}) + d_{n+1}$$

- On commence donc pas **développer le carré** de $a_{n+1}(x)$:

$$\begin{aligned} (q_{n+1}(x) - q(x))^2 &= (q_n(x) - q(x))^2 + \gamma_{n+1}^2 \left[(1 - 2\alpha) \mathbf{1}_{Y_{n+1} \leq q_n(x)} + \alpha^2 \right] \mathbf{1}_{X_{n+1} \in kNN_n(x)} \\ &\quad - 2\gamma_{n+1}(q_n(x) - q(x)) \left(\mathbf{1}_{Y_{n+1} \leq q_n(x)} - \alpha \right) \mathbf{1}_{X_{n+1} \in kNN_n(x)} \end{aligned}$$

- On prend **l'espérance conditionnelle** et on utilise la formule de Bayes

$$\begin{aligned} \mathbb{E}_n \left((q_{n+1}(x) - q(x))^2 \right) &\leq \mathbb{E}_n \left((q_n(x) - q(x))^2 \right) + \gamma_{n+1}^2 P_n \\ &\quad - 2\gamma_{n+1} (q_n(x) - \theta^*(x)) P_n \left[F_{Y^{B_n(x)}}(q_n(x)) - F_{Y^x}(q(x)) \right] \end{aligned}$$

Esquisse de preuve

- $F_{Y^{B_n(x)}}(q_n(x)) - F_{Y^x}(q_n(x))$, erreur **de biais**.

Par **A1**,

$$|F_{Y^{B_n^{k_{n+1}}(x)}(x)}(q_n(x)) - F_{Y^x}(q_n(x))| \leq M(x) \|X - x\|_{(k_{n+1}, n)}$$

et par **A3**, $|q_n(x) - q(x)| \leq \sqrt{C_1}$ donc

$$\begin{aligned} & -2\gamma_{n+1}(q_n(x) - q(x))P_n \left[F_{Y^{B_n^{k_{n+1}}(x)}(x)}(q_n(x)) - F_{Y^x}(q_n(x)) \right] \\ & \leq 2\gamma_{n+1}\sqrt{R}M(x)P_n \|X - x\|_{(k_n, n)} \end{aligned}$$

- $F_{Y^x}(q_n(x)) - F_{Y^x}(q(x))$, erreur **d'apprentissage**.

Par **A4** nous avons

$$(q_n - q) [F_{Y^x}(q_n(x)) - F_{Y^x}(q(x))] \geq C_2(x, \alpha)(x) [q_n(x) - q(x)]^2.$$

donc

$$\begin{aligned} & -2\gamma_{n+1}(q_n(x) - q(x))P_n [F_{Y^x}(q(x)) - F_{Y^x}(q(x))] \\ & \leq -2\gamma_{n+1} C_2(x, \alpha)(q_n(x) - q(x))^2 P_n \end{aligned}$$

Esquisse de preuve

- Finalement,

$$\begin{aligned}\mathbb{E}_n (q_{n+1}(x) - q(x))^2 &\leq (q_n(x) - q(x))^2 - 2\gamma_{n+1} C_2(x, \alpha)(q_n(x) - q(x))^2 P_n \\ &\quad + \gamma_{n+1}^2 P_n + 2\gamma_{n+1} M(x) \sqrt{C_1} \|X - x\|_{(k_{n+1}, n)} P_n.\end{aligned}$$

- On prend l'espérance

$$\begin{aligned}a_{n+1}(x) &\leq a_n(x) - 2\gamma_{n+1} D_{code}(x) \mathbb{E} [(\theta_n(x) - \theta^*(x))^2 P_n] \\ &\quad + \gamma_{n+1}^2 \mathbb{E}(P_n) + 2\gamma_{n+1} M \sqrt{R} \mathbb{E}(\|X - x\|_{(k, n)} P_n).\end{aligned}$$

Esquisse de preuve

- $\mathbb{E}(P_n) = \frac{k_{n+1}}{n+1}$ car

$$P_n = \mathbb{P}(\|X - x\| \leq \|X - x\|_{(k_{n+1}, n)}) = F_{\|X - x\|}(\|X - x\|_{(k_{n+1}, n)}) \sim \beta(k_{n+1}, n - k_{n+1} + 1)$$

- $\mathbb{E}(\|X - x\|_{(k_n, n)} P_n) \leq C_3(d) \left(\frac{k_{n+1}}{n+1} \right)^{1 + \frac{1}{d}}$ par l'inégalité de Bernstein.

Esquisse de preuve

- $\mathbb{E}(P_n(q_n(x) - q(x))^2) = ?$

Solution : On va utiliser une borne sur cette espérance. Pour cela, on doit considérer $b_n(x) = \mathbb{E}((q_n(x) - q^*(x))^2 \mathbf{1}_{P_n > \epsilon_{n+1}})$ avec $\epsilon_n = \frac{1}{n^\epsilon}$.

On peut faire les mêmes raisonnements sur cette quantité pour trouver :

$$b_{n+1}(x) \leq b_n(x)(1 - c'_{n+1}) + d'_{n+1}$$

Pour se ramener au même type d'inégalité sur $a_n(x)$, on gère le terme $\overline{b_n(x)} := \mathbb{E}((q_n(x) - q(x))^2 \mathbf{1}_{P_n \leq \epsilon_n})$ avec des inégalités de concentration.

Comment estimer le superquantile ?

- Par définition, le superquantile est lié au quantile :

$$Q_\alpha(Z) = \mathbb{E} \left(\frac{Z \mathbf{1}_{Z > q_\alpha(Z)}}{1 - \alpha} \right).$$

- L'estimateur empirique de la moyenne peut s'écrire sous forme d'un algorithme stochastique :

$$\overline{Z}_{n+1} = \overline{Z}_n + \frac{1}{n+1} (Z_{n+1} - \overline{Z}_n).$$

Double algorithme (Bardou et al. - 2009)

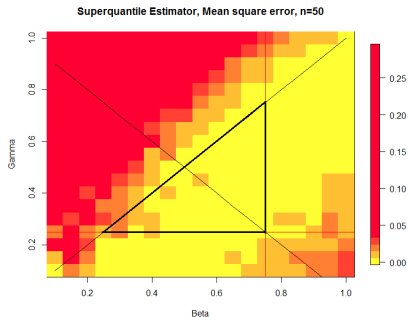
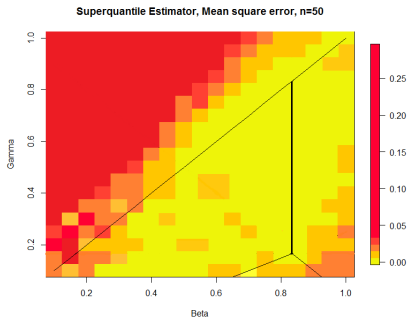
- On peut donc estimer à la fois quantile et superquantile par le double algorithme :

$$\begin{cases} (q_0, Q_0) \in \mathbb{R}^2 \\ q_{n+1} = q_n - \gamma_{n+1}^q (\mathbf{1}_{Y_{n+1} \leq q_n} - \alpha) \\ Q_{n+1} = Q_n + \gamma_{n+1}^Q (Y_{n+1} \mathbf{1}_{Y_{n+1} > q_n} - Q_n) \end{cases}$$

- De la même manière, on peut estimer le quantile et le superquantile de la sortie d'un code stochastique par :

$$\begin{cases} (q_0(x), Q_0(x)) \in \mathbb{R}^2 \\ q_{n+1}(x) = q_n(x) - \gamma_{n+1}^q (\mathbf{1}_{Y_{n+1} > q_n(x)} - \alpha) \mathbf{1}_{X_{n+1} \in kNN_{n+1}^q(x)} \\ Q_{n+1}(x) = Q_n(x) + \gamma_{n+1}^Q (Y_{n+1} \mathbf{1}_{Y_{n+1} > q_n(x)} - Q_n(x)) \mathbf{1}_{X_{n+1} \in kNN_{n+1}^Q(x)} \end{cases}$$

Paramètres optimaux



Fonctions tests

La fonction de Branin :

$$g(\mathbf{x}) = \left(\bar{x}_2 - \frac{5.1\bar{x}_1^2}{4\pi^2} + \frac{5\bar{x}_1}{\pi} - 6 \right)^2 + \left(10 - \frac{10}{8\pi} \right) \cos(\bar{x}_1) + 10$$

avec $\bar{x}_1 = 15 \times x_1 - 5$, $\bar{x}_2 = 15 \times x_2$.

Fonctions tests

La fonction de Hartman :

$$g(\mathbf{x}) = \frac{-1}{1.94} \left[2.58 + \sum_{i=1}^4 C_i \exp \left(- \sum_{j=1}^4 a_{ji} (x_j - p_{ji})^2 \right) \right],$$

avec

$$\mathbf{C} = \begin{bmatrix} 1.0 \\ 1.2 \\ 3.0 \\ 3.2 \end{bmatrix}, \quad \mathbf{a} = \begin{bmatrix} 10.00 & 0.05 & 3.00 & 17.00 \\ 3.00 & 10.00 & 3.50 & 8.00 \\ 17.00 & 17.00 & 1.70 & 0.05 \\ 3.50 & 0.10 & 10.00 & 10.00 \end{bmatrix},$$

$$\mathbf{p} = \begin{bmatrix} 0.1312 & 0.2329 & 0.2348 & 0.4047 \\ 0.1696 & 0.4135 & 0.1451 & 0.8828 \\ 0.5569 & 0.8307 & 0.3522 & 0.8732 \\ 0.0124 & 0.3736 & 0.2883 & 0.5743 \end{bmatrix}.$$

Functions tests

La fonction de Ackley :

$$g(\mathbf{x}) = 20 + \exp(1) - 20 \exp \left(-0.2 \sqrt{\frac{1}{4} \sum_{i=1}^4 x_i^2} \right) - \exp \left[\frac{1}{4} \sum_{i=1}^4 \cos(2\pi x_i) \right]$$