



Estimation de quantiles de codes Application thermohydraulique

Projet CEMRACS 2006 proposé par

B. looss, N. Devictor (CEA Cadarache/DER)

A. de Crecy, P. Bazin (CEA Grenoble/DER)

traité par

J. Garnier (Université Paris VII)

C. Cannamela, B. looss (CEA Cadarache)

Le CEMRACS 2006 ???

Centre d'Été Mathématique de Recherche Avancée en Calcul Scientifique



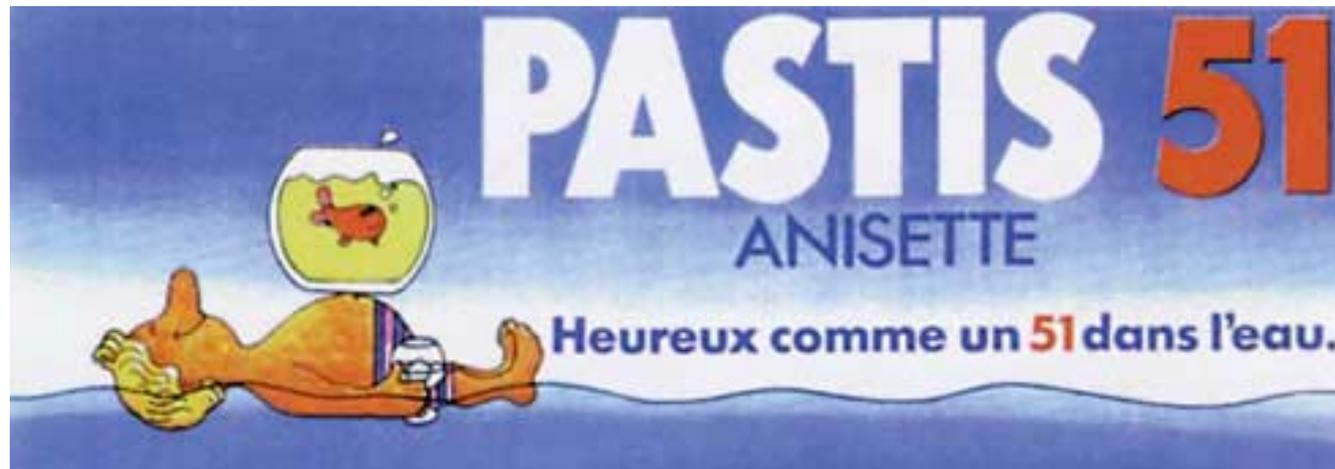
Le CEMRACS 2006, c'est d'abord ça !

Centre d'Été Mathématique de Recherche Avancée en Calcul Scientifique



Mais c'est aussi ça !!

cea



Et enfin c'est ça : « Estimation de quantiles !!! »

- Déterminer $\hat{Y}_{\alpha,n}$ estimateur de y_α / $P(Y < y_\alpha) = \alpha$

à partir d'un n-échantillon (Y_1, \dots, Y_n) ;

$Y = f(X)$ est la réponse d'un code où $X = (X_1, \dots, X_p)$ est aléatoire.



Et enfin c'est ça : « Estimation de quantiles !!! »

- Déterminer $\hat{Y}_{\alpha,n}$ estimateur de y_α / $P(Y < y_\alpha) = \alpha$
à partir d'un n-échantillon (Y_1, \dots, Y_n) ;

$Y = f(X)$ est la réponse d'un code où $X = (X_1, \dots, X_p)$ est aléatoire.



- Pour assurer le conservatisme, on introduit parfois un niveau de confiance β sur la probabilité : $P(\hat{Y}_{\alpha,n} \geq y_\alpha) \geq \beta$

Et enfin c'est ça : « Estimation de quantiles !!! »

- Déterminer $\hat{Y}_{\alpha,n}$ estimateur de y_α / $P(Y < y_\alpha) = \alpha$
à partir d'un n-échantillon (Y_1, \dots, Y_n) ;
 $Y = f(X)$ est la réponse d'un code où $X = (X_1, \dots, X_p)$ est aléatoire.



- Pour assurer le conservatisme, on introduit parfois un niveau de confiance β sur la probabilité : $P(\hat{Y}_{\alpha,n} \geq y_\alpha) \geq \beta$
- **Objectif du projet** : estimation d'un quantile à $\alpha=95\%$ avec $n=200$.
- **Solution simple** : quantile empirique par simulations Monte-Carlo.
Si $\alpha=95\%$, cela nécessite plusieurs centaines de calculs.
- **Solution classique dans le nucléaire** : formule de Wilks (statistiques d'ordre) : requiert peu de calculs mais dispersion forte (> dispersion de l'estimateur empirique).
- **Solution qui ne marche pas** : quantile empirique à l'aide d'une surface de réponse : lisse les résultats de code et donc sous-estime le quantile.

Estimation de quantiles : le projet BEMUSE

- Benchmark OCDE/NEA : **analyses d'incertitudes et de sensibilité sur les calculs thermohydrauliques de sûreté des réacteurs nucléaires.**



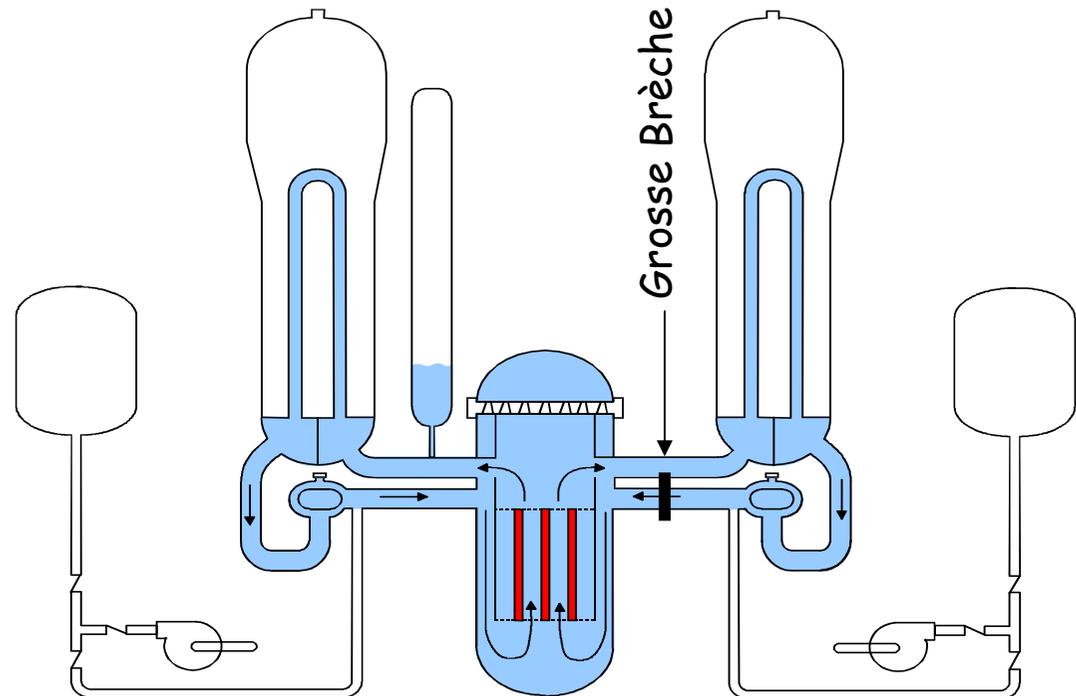
- Piloté par le CEA Grenoble/LDAS (A. De Crecy, P. Bazin), participation de : IRSN, GRS, KAERI, JNES, ...

Scénario :

Accident de Perte de Réfrigérant Primaire due à une Grosse Brèche

Étude CEA :

Simulation avec le logiciel CATHARE



Sources d'incertitudes

Identification de 53 paramètres d'entrée incertains



Types des paramètres :

- Modèles physiques de CATHARE (25) :
jugements d'expert, méthode inverse (CIRCE).
- Conditions initiales et aux limites (13) :
expériences, jugements d'expert.
- Propriétés des matériaux (3) :
données, jugements d'expert.
- Modélisation géométrique (11) :
expériences, jugements d'expert.

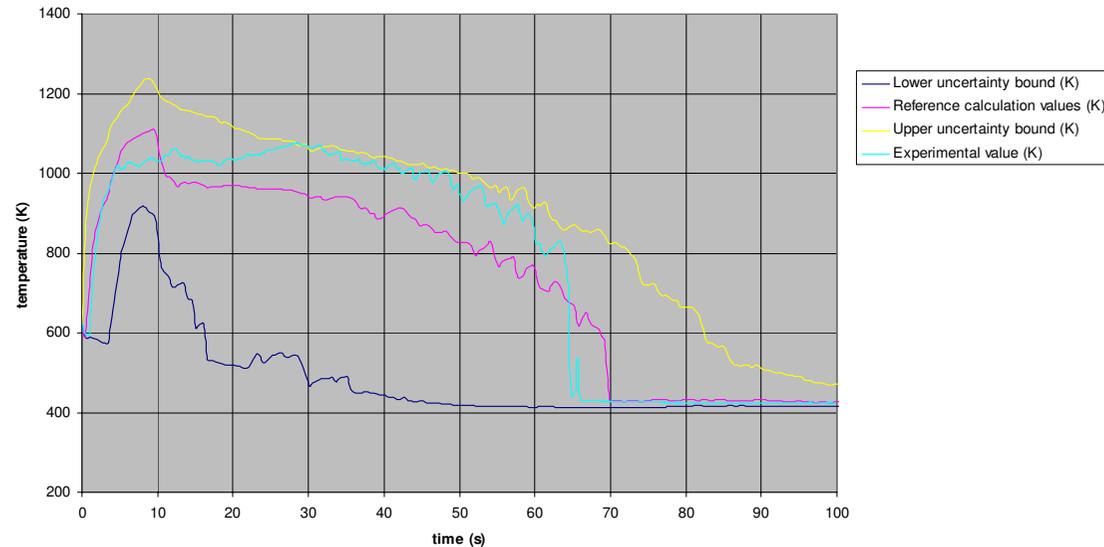
Objectifs de l'analyse d'incertitudes

Exemples de sorties étudiées :

- Évolution de la température maximale de gaine dans le cœur,
- valeur du 1er PCT (pic de température de gaine obtenue exp. vers 12 s).



CEA: maximum cladding temperature



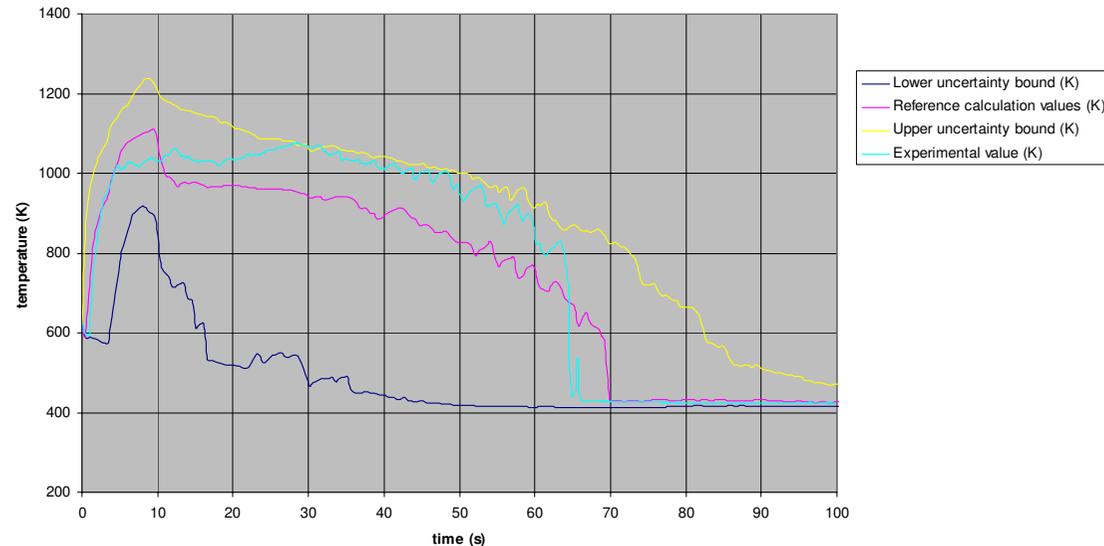
Objectifs de l'analyse d'incertitudes

Exemples de sorties étudiées :

- Évolution de la température maximale de gaine dans le cœur,
- valeur du 1er PCT (pic de température de gaine obtenue exp. vers 12 s).



CEA: maximum cladding temperature



Objectif important de BEMUSE :

Vérifier que le calcul de ce scénario respecte certains critères.

Par ex : $T_{\text{gaine_max}} + \text{incertitude} < 1477 \text{ K}$

Étude CEMRACS : obtention du quantile à 95% du 1er PCT.

Plan

1. Estimateurs classiques : empirique, Wilks



2. Variables de contrôle et stratification contrôlée

3. Exemples analytiques

4. Application à CATHARE

Estimation de quantile : l'estimateur empirique

Échantillon $(Y_1, \dots, Y_n) \rightarrow$ estimateur $\hat{Y}_{\alpha,n}$ du α -quantile $y_\alpha / P(Y \leq y_\alpha) = \alpha$



Échantillon ordonné : $(Y_{(1)}, \dots, Y_{(n)}) / Y_{(1)} \leq \dots \leq Y_{(n)}$

Estimateur empirique : $\hat{Y}_{\alpha,n} = Y_{(\lfloor \alpha n \rfloor + 1)}$

Résultats asymptotiques (Bahadur, 1966), avec p la densité de Y :

$$E(\hat{Y}_{\alpha,n}) = y_\alpha - \frac{\alpha(1-\alpha)p'(y_\alpha)}{2(n+2)p^3(y_\alpha)} + O\left(\frac{1}{n^2}\right)$$

$$\text{Var}(\hat{Y}_{\alpha,n}) = \frac{\alpha(1-\alpha)}{(n+2)p^2(y_\alpha)} + O\left(\frac{1}{n^2}\right)$$

$$\sqrt{n}(\hat{Y}_{\alpha,n} - y_\alpha) \xrightarrow{n \rightarrow \infty} \mathbf{N} \left(0, \frac{\alpha(1-\alpha)}{p^2(y_\alpha)} \right)$$

Estimation de quantile : l'estimateur de Wilks

Ainsi, pour n grand : $P(\hat{Y}_{\alpha,n} \geq \gamma_\alpha) \approx 0.5$.



Pour augmenter cette proba., on peut prendre $\hat{Y}_{\alpha,n} = Y_{(\lfloor \alpha n \rfloor + s)}$.

Comment choisir s / $P(Y_{(\lfloor \alpha n \rfloor + s)} \geq \gamma_\alpha) \geq \beta$?

Estimation de quantile : l'estimateur de Wilks

Ainsi, pour n grand : $P(\hat{Y}_{\alpha,n} \geq \gamma_\alpha) \approx 0.5$.



Pour augmenter cette proba., on peut prendre $\hat{Y}_{\alpha,n} = Y_{(\lfloor \alpha n \rfloor + s)}$.

Comment choisir s / $P(Y_{(\lfloor \alpha n \rfloor + s)} \geq \gamma_\alpha) \geq \beta$?

Proposition : Le nb de dépassements d'un seuil y par une suite de n v.a. i.i.d suit une loi binomiale $B(n, F(y))$ où F est la fct de répartition des Y_i .

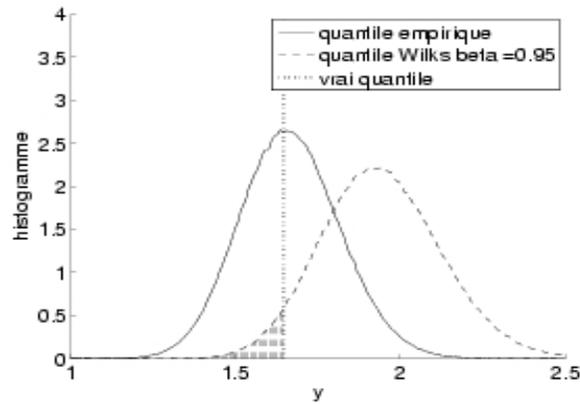
→ **Formule de Wilks qui lie n et l'ordre r** (rang dans la suite décroissante $(Y_{(n)}, \dots, Y_{(n-r+1)}, \dots, Y_{(1)})$), où $Y_{(n-r+1)}$ est l'estimateur du quantile :

$$1 - G(\alpha) = 1 - \sum_{n-r+1}^n C_n^i \alpha^i (1-\alpha)^{(n-i)} \geq \beta \quad \text{qui devient pour } r=1 : 1 - \alpha^n \geq \beta$$

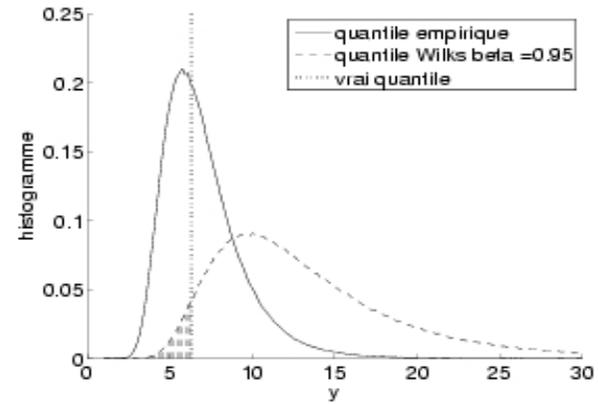
Pour $\alpha = \beta = 95\%$, on a les couples suivants :

$(r = 1, n = 59)$; $(r = 2, n = 93)$; $(r = 3, n = 124)$... $(r = 39, n = 991)$.

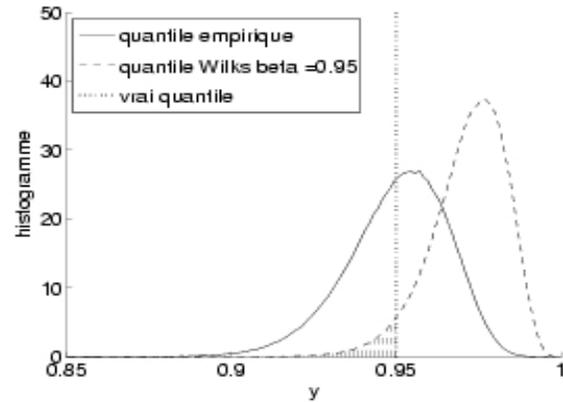
Comparaisons estimateur empirique / Wilks (n=200)



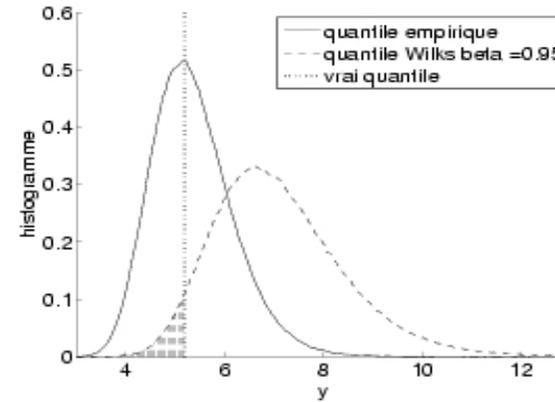
X N (0,1)



X Cauchy



X U (0,1)



X log-normal

Plan

1. Estimateurs classiques : empirique, Wilks



2. Variables de contrôle et stratification contrôlée

3. Exemples analytiques

4. Application à CATHARE

Variable de contrôle

On revient au projet CEMRACS :

On a $Y=f(X)$, où X est un vecteur de paramètres aléatoires.

On suppose que l'on a aussi une surface de réponse $Z=f_r(X)$,
modèle qui approche approximativement $f(X)$.

f est coûteux mais f_r est très peu coûteux (on connaît $z_\alpha \forall \alpha$).

Idée : utiliser Z pour réduire la variance de l'estimateur du quantile



Variable de contrôle

On revient au projet CEMRACS :

On a $Y=f(X)$, où X est un vecteur de paramètres aléatoires.

On suppose que l'on a aussi une surface de réponse $Z=f_r(X)$,
modèle qui approche approximativement $f(X)$.

f est coûteux mais f_r est très peu coûteux (on connaît $z_\alpha \forall \alpha$).

Idée : utiliser Z pour réduire la variance de l'estimateur du quantile

On pose

$$p_{00}(y) = P(Z \leq z_\alpha, Y \leq y) \quad p_{01}(y) = P(Z \leq z_\alpha, Y > y)$$

$$p_{10}(y) = P(Z > z_\alpha, Y \leq y) \quad p_{11}(y) = P(Z > z_\alpha, Y > y)$$

On a les relations $p_{00}(y) + p_{01}(y) = P(Z \leq z_\alpha) = \alpha$

et $p_{10}(y) + p_{11}(y) = P(Z > z_\alpha) = 1 - \alpha$

Alors, on peut estimer $F(y) = P(Y \leq y) = p_{00}(y) + p_{10}(y)$

Fonction de répartition par variable de contrôle

On a un n-échantillon de v.a. i.i.d. $((Y_1, Z_1), \dots, (Y_n, Z_n))$, on note

$$N_{00}(y) = \text{Card} \{ i=1, \dots, n / Z_i \leq z_\alpha, Y_i \leq y \}$$

$$N_{01}(y) = \text{Card} \{ i=1, \dots, n / Z_i \leq z_\alpha, Y_i > y \}$$

$$N_{10}(y) = \text{Card} \{ i=1, \dots, n / Z_i > z_\alpha, Y_i \leq y \}$$

$$N_{11}(y) = \text{Card} \{ i=1, \dots, n / Z_i > z_\alpha, Y_i > y \}$$

$$N_0 = \text{Card} \{ i=1, \dots, n / Z_i \leq z_\alpha \}$$

$$N_1 = \text{Card} \{ i=1, \dots, n / Z_i > z_\alpha \}$$



Fonction de répartition par variable de contrôle

On a un n-échantillon de v.a. i.i.d. $((Y_1, Z_1), \dots, (Y_n, Z_n))$, on note

$$N_{00}(y) = \text{Card} \{ i=1, \dots, n / Z_i \leq z_\alpha, Y_i \leq y \}$$

$$N_{01}(y) = \text{Card} \{ i=1, \dots, n / Z_i \leq z_\alpha, Y_i > y \}$$

$$N_{10}(y) = \text{Card} \{ i=1, \dots, n / Z_i > z_\alpha, Y_i \leq y \}$$

$$N_{11}(y) = \text{Card} \{ i=1, \dots, n / Z_i > z_\alpha, Y_i > y \}$$

$$N_0 = \text{Card} \{ i=1, \dots, n / Z_i \leq z_\alpha \}$$

$$N_1 = \text{Card} \{ i=1, \dots, n / Z_i > z_\alpha \}$$

La loi de $(N_{00}, N_{01}, N_{10}, N_{11})$ est multinomiale, on a alors les estimations par max. de vraisemblance :

$$\hat{p}_{00}(y) = \frac{\alpha N_{00}(y)}{N_0} \quad \text{et} \quad \hat{p}_{10}(y) = \frac{(1-\alpha) N_{10}(y)}{N_1}$$

Et l'estimateur de $F(y)$:

$$\hat{F}(y) = \frac{\alpha N_{00}(y)}{N_0} + \frac{(1-\alpha) N_{10}(y)}{N_1}$$



Estimation de quantile par variable de contrôle

L'estimateur de $F(y)$ peut s'écrire $\hat{F}(y) = \sum_{j=1}^n W_j 1_{Y_j \leq y}$

avec $W_j = \frac{\alpha}{N_0} 1_{Z_j \leq z_\alpha} + \frac{(1-\alpha)}{N_1} 1_{Z_j > z_\alpha}$

On considère l'échantillon ordonné $(Y_{(1)}, \dots, Y_{(n)})$, on en déduit l'estimateur

$$\hat{Y}_{\alpha, n} = Y_{(K)} \quad \text{avec} \quad K = \inf \left\{ j / \sum_{i=1}^j W_{(i)} > \alpha \right\}$$

Estimation de quantile par variable de contrôle

L'estimateur de $F(y)$ peut s'écrire $\hat{F}(y) = \sum_{j=1}^n W_j 1_{Y_j \leq y}$

avec $W_j = \frac{\alpha}{N_0} 1_{Z_j \leq z_\alpha} + \frac{(1-\alpha)}{N_1} 1_{Z_j > z_\alpha}$

On considère l'échantillon ordonné $(Y_{(1)}, \dots, Y_{(n)})$, on en déduit l'estimateur

$\hat{Y}_{\alpha,n} = Y_{(K)}$ avec $K = \inf \left\{ j / \sum_{i=1}^j W_{(i)} > \alpha \right\}$

Asymptotique : $\sqrt{n}(\hat{Y}_{\alpha,n} - y_\alpha) \xrightarrow{n \rightarrow \infty} N \left(0, \frac{\alpha(1-\alpha)}{p^2(y_\alpha)} (1 - \rho_I^2) \right)$

où ρ_I est le coefficient de corrélation entre $1_{Y \leq y_\alpha}$ et $1_{Z \leq z_\alpha}$.

La variable de contrôle a permis de réduire la variance de l'estimateur d'un facteur $1 - \rho_I^2$ (que l'on peut estimer).

Remarques sur la variable de contrôle

- On n'a pas utilisé le fait que la surface de réponse $Z=f_r(X)$ est très peu coûteuse en temps de calcul.



- En général, l'utilisation directe de Z pour estimer un quantile élevé n'est pas efficace, car Z ajuste $Y=f(X)$ pour des réalisations typiques et non pour des réalisations exceptionnelles.

- **Idée : utiliser f_r pour trouver des échantillons de X dans des zones où on s'attend à ce qu'il se passe quelque chose.**

Remarques sur la variable de contrôle

- On n'a pas utilisé le fait que la surface de réponse $Z=f_r(X)$ est très peu coûteuse en temps de calcul.



- En général, l'utilisation directe de Z pour estimer un quantile élevé n'est pas efficace, car Z ajuste $Y=f(X)$ pour des réalisations typiques et non pour des réalisations exceptionnelles.

- **Idée : utiliser f_r pour trouver des échantillons de X dans des zones où on s'attend à ce qu'il se passe quelque chose.**

→ **Stratification contrôlée :**

- Diviser le domaine de variation de Z en m strates (intervalles) I_1, \dots, I_m
- Contrôler les réalisations de $X / Z=f_r(X) \in I_j$
- Favoriser les strates où f_r prédit de grandes valeurs en leur allouant plus de réalisations pour le calcul de $Y=f(X)$

Fonction de répartition par stratification contrôlée

- On considère $m+1$ niveaux : $0 = \alpha_0 < \alpha_1 < \dots < \alpha_m = 1$.

Quantiles correspondant : $-\infty < z_{\alpha_1} < \dots < z_{\alpha_m} = \infty$ (définissent m strates).

On prend une séquence d'entiers N_1, \dots, N_m / $\sum_{j=1}^m N_j = n$

- La loi des probabilités totales donnent

$$F(y) = P(Y \leq y) = \sum_{j=1}^m P(Y \leq y \mid Z \in]z_{\alpha_{j-1}}, z_{\alpha_j}]) (\alpha_j - \alpha_{j-1}) = \sum_{j=1}^m p_j(y) (\alpha_j - \alpha_{j-1})$$



Fonction de répartition par stratification contrôlée

- On considère $m+1$ niveaux : $0 = \alpha_0 < \alpha_1 < \dots < \alpha_m = 1$.

Quantiles correspondant : $-\infty < z_{\alpha_1} < \dots < z_{\alpha_m} = \infty$ (définissent m strates).

On prend une séquence d'entiers N_1, \dots, N_m / $\sum_{j=1}^m N_j = n$

- La loi des probabilités totales donnent

$$F(y) = P(Y \leq y) = \sum_{j=1}^m P(Y \leq y \mid Z \in]z_{\alpha_{j-1}}, z_{\alpha_j}]) (\alpha_j - \alpha_{j-1}) = \sum_{j=1}^m p_j(y) (\alpha_j - \alpha_{j-1})$$

- Estimation de $p_j(y)$ ($j=1, \dots, m$) :

- On génère $(X_i^{(j)})_{i=1 \dots N_j}$ / $Z_i^{(j)} = f_r(X_i^{(j)}) \in]z_{\alpha_{j-1}}, z_{\alpha_j}]$;

- On calcule $Y_i^{(j)} = f(X_i^{(j)})$ pour $i=1, \dots, N_j$;

- On trouve

$$\hat{p}_j(y) = \sum_{i=1}^{N_j} 1_{Y_i^{(j)} \leq y} \quad \text{et ainsi} \quad \hat{F}(y).$$

Remarques



• Si Z et Y sont indépendantes, $\text{Var}[\hat{F}(y)] = \text{Var}[\hat{F}(y)]|_{\text{sans contrôle}}$

• Si Z est une fonction croissante de Y (contrôle total), $\alpha_j = j/m$ et $N_j = n/m$, la variance est réduite d'un facteur $1/m$.

• Si on vise la queue de $F(y)$, on a intérêt à y mettre plus de points.

*Par ex $\alpha = 0.95$: $m = 4$, $N_j = n/4$ ($j = 1, \dots, 4$), $\alpha_1 = 0.5$, $\alpha_2 = 0.9$, $\alpha_3 = 0.95$
($n/2$ points centrés autour du quantile cherché).*

Estimation de quantile par stratification contrôlée



- **Estimateur du quantile :**

$$\hat{Y}_{\alpha,n} = \inf \left\{ y / \hat{F}(y) > \alpha \right\}$$

- Asymptotique (avec $\beta_j = N_j / n$) :

$$\sqrt{n}(\hat{Y}_{\alpha,n} - y_\alpha) \xrightarrow{n \rightarrow \infty} N \left(0, \frac{\sum_{j=1}^m \frac{(\alpha_j - \alpha_{j-1})^2}{\beta_j} [p_j(y_\alpha) - p_j(y_\alpha)^2]}{p^2(y_\alpha)} \right)$$

➡ il faut mettre + de points dans les zones intéressantes

Plan

1. Estimateurs classiques : empirique, Wilks



2. Variables de contrôle et stratification contrôlée

3. Exemples analytiques

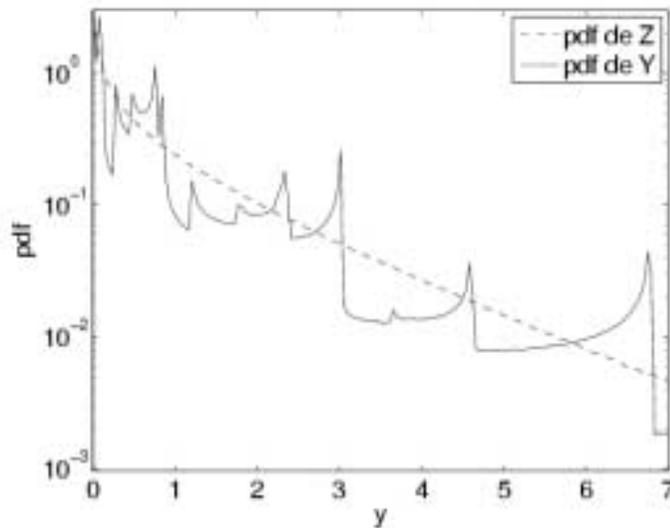
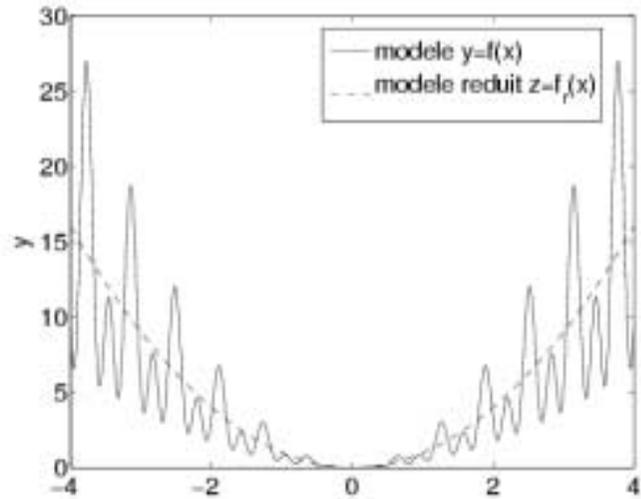
4. Application à CATHARE

Exemple – 1 paramètre - $X \sim N(0,1)$

$$Y=f(x) = 0.95x^2 [1 + 0.5\cos(10x) + 0.5\cos(20x)]$$

$$Z=f_r(x)=x^2$$

ce



Exemple – 1 paramètre - $X \sim N(0,1)$

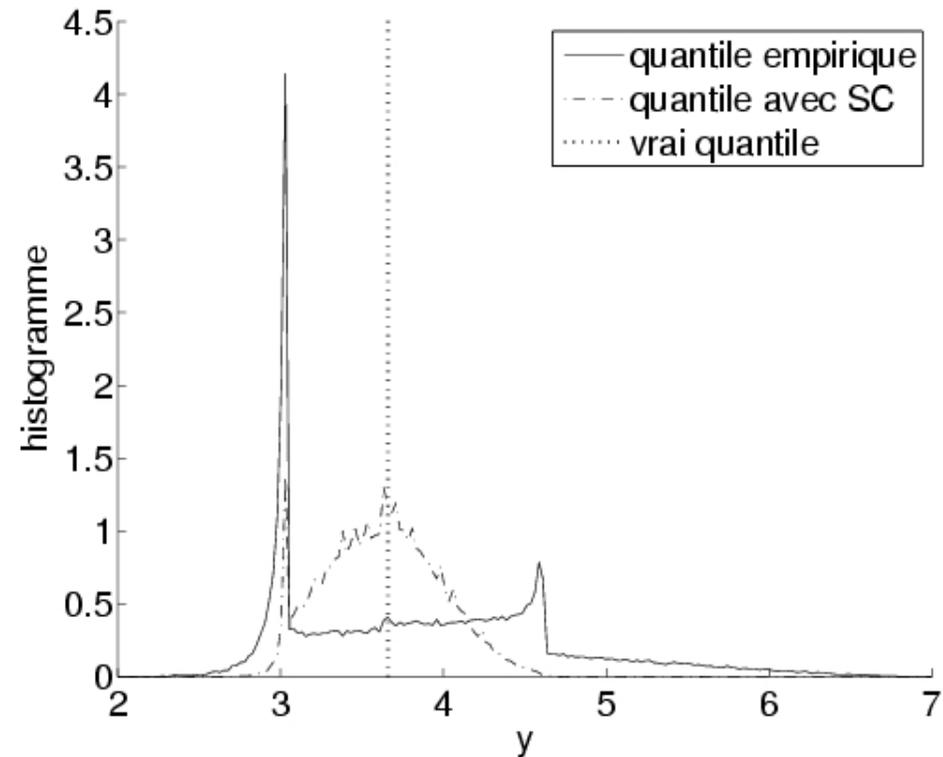
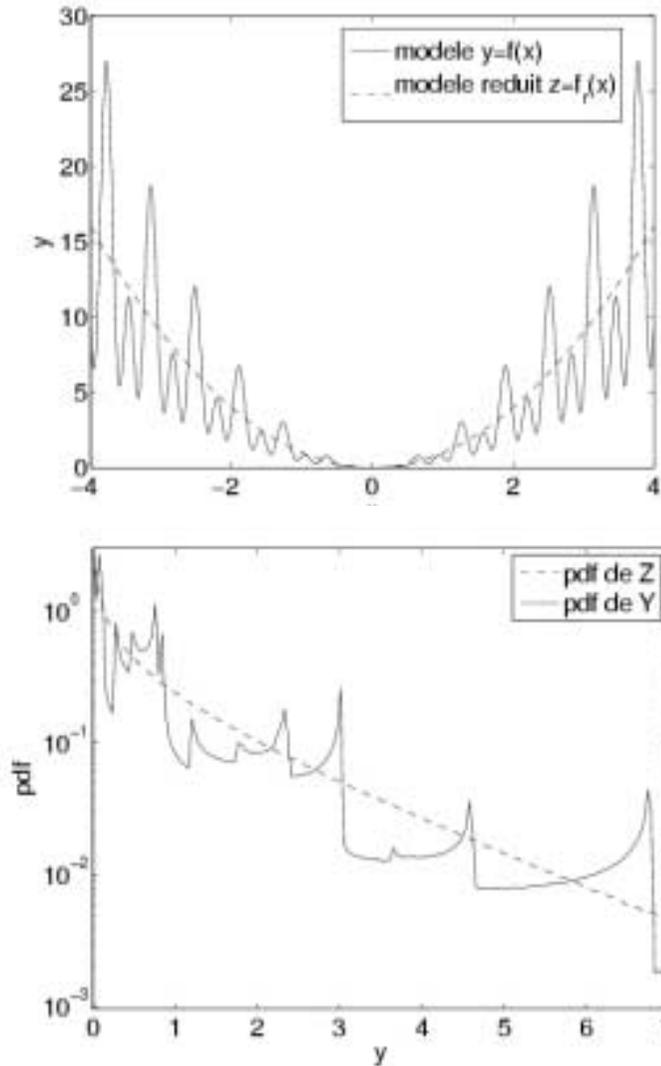
$$Y=f(x) = 0.95x^2 [1 + 0.5\cos(10x) + 0.5\cos(20x)]$$

$$Z=f_r(x)=x^2$$

ce

$\alpha=0.95$, $n=200$, $m=4$ strates

Histogrammes avec
10000 expériences



Fonction d'Ishigami – 3 paramètres

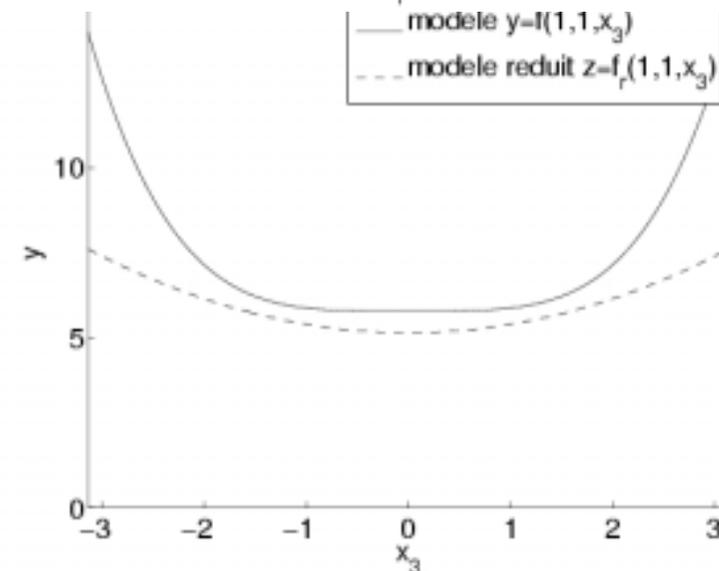
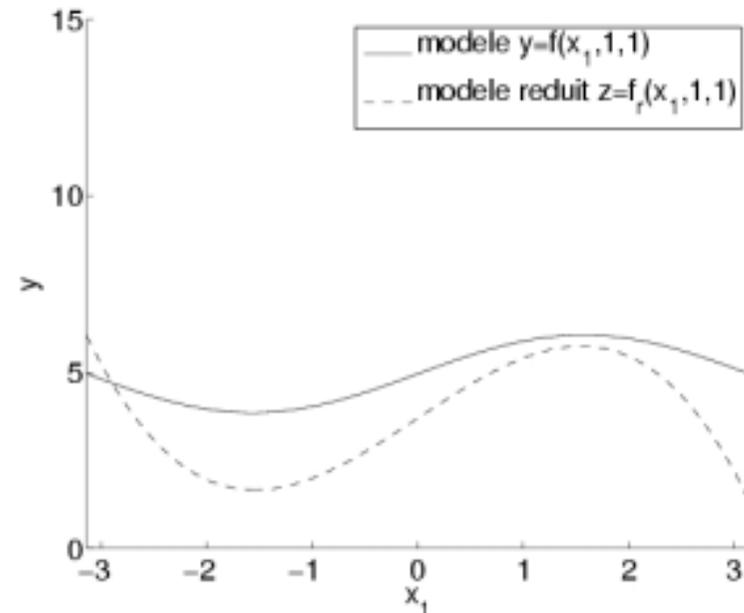
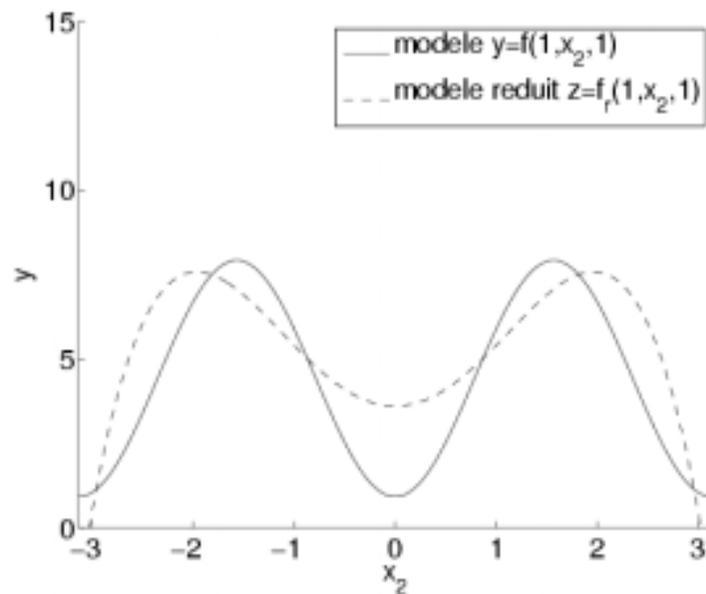
$$Y = f(x_1, x_2, x_3) = \sin(x_1) + 7\sin(x_2)^2 + 0.1 x_3^4 \sin(x_1)$$

CEA

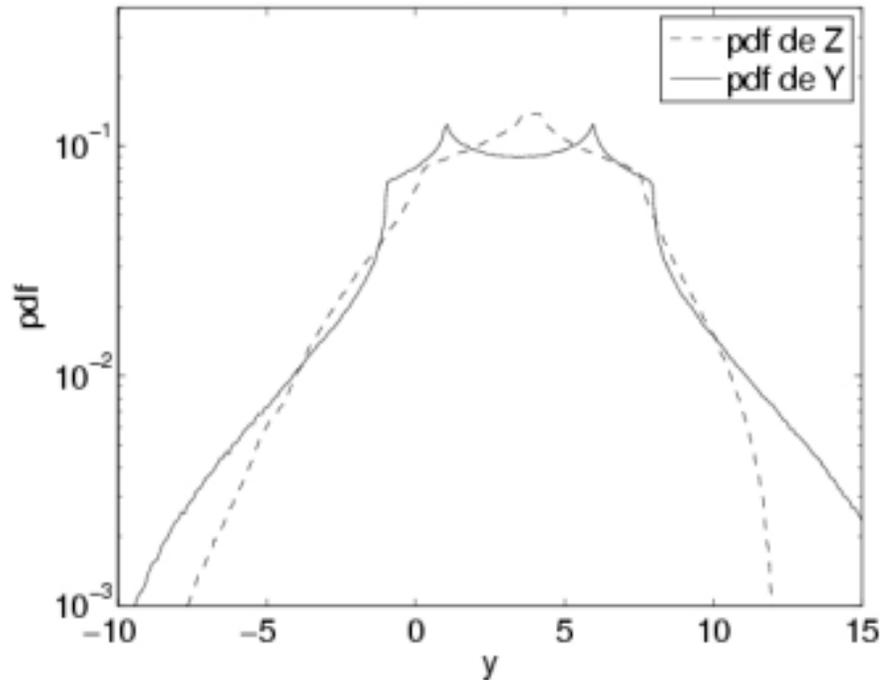
$$Z = f_r(x_1, x_2, x_3) = 1.66x_1 + 1.98x_2^2 - 0.27x_1^3 - 0.26x_2^4 + 0.25x_3^2 x_1$$

$X_1, X_2, X_3 \sim U[-\pi, \pi]$

En prédiction : $R^2=0.75$; $\rho=0.86$



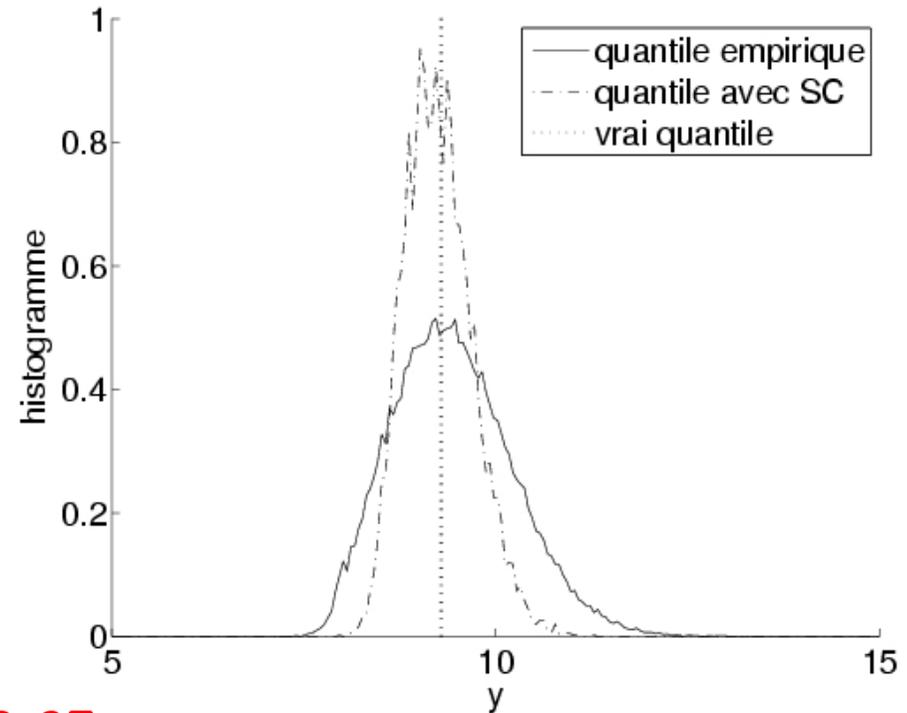
Fonction d'Ishigami



$\alpha=0.95$

α -quantile de Y : 9.30

α -quantile de Z : 8.51



n = 200

4 strates

Histogrammes avec
10000 expériences

Fonction de Morris – 10 paramètres

Fonction à 10 paramètres d'entrée ($\sim U[0,1]$) utilisée en analyse de sensibilité (effets négligeables, linéaires, non linéaires, interactions).

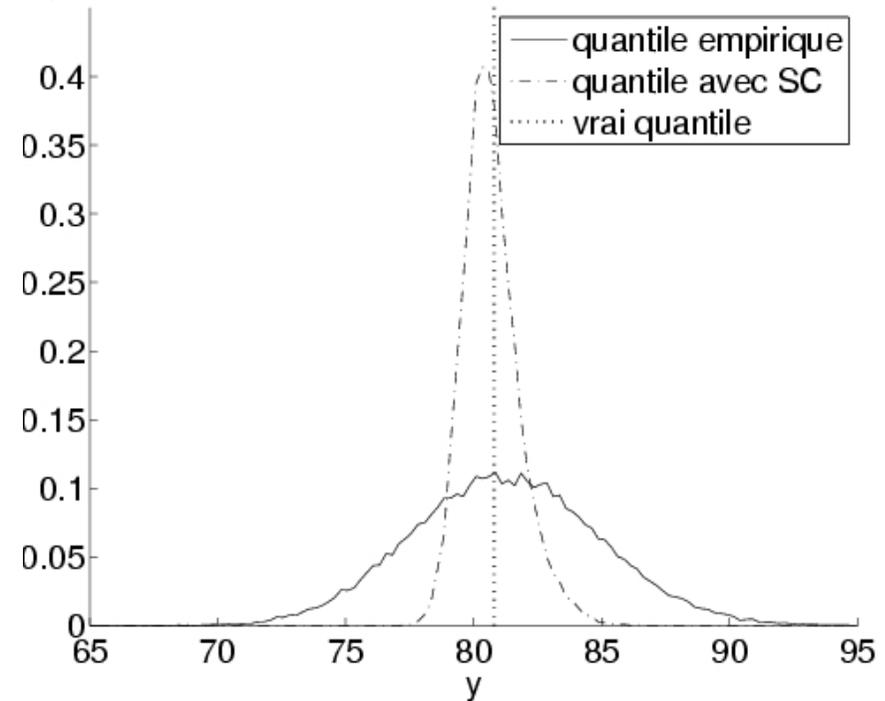
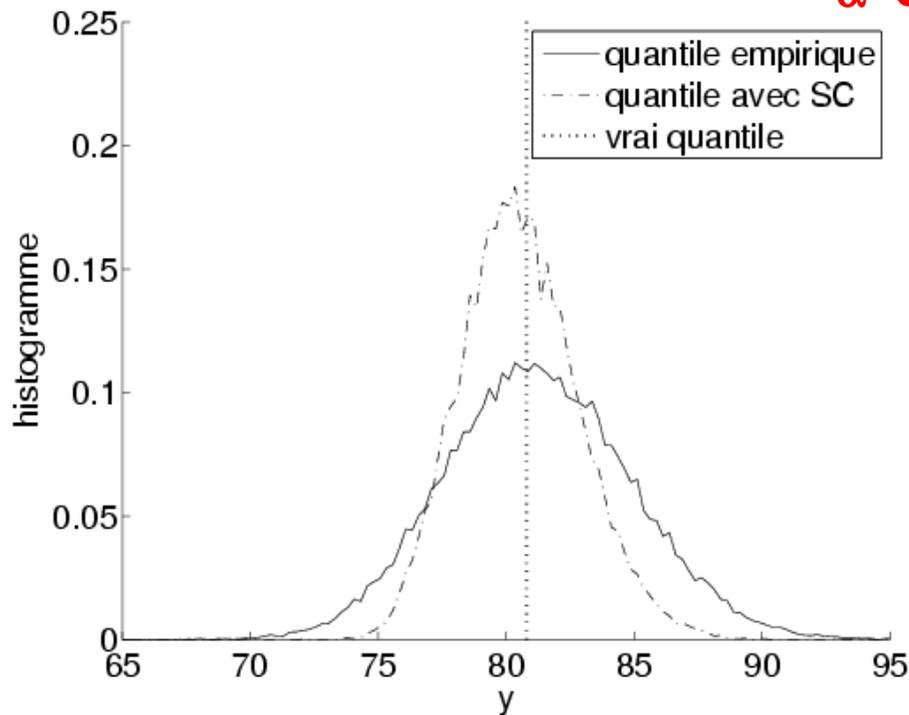


Deux surfaces de réponse polynomiale (ordre 2) :

- ($R^2=0.7$; $\rho=0.86$)

- ($R^2=0.9$; $\rho=0.95$)

$\alpha=0.95$, $n=200$



Plan

1. Estimateurs classiques : empirique, Wilks



2. Variables de contrôle et stratification contrôlée

3. Exemples analytiques

4. Application à CATHARE

Application à CATHARE

- On suppose que l'on ne connaît rien du modèle



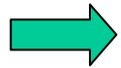
- Pour la stratification contrôlée, il faut construire une surface de réponse avec le moins de calculs possible (problème : 53 paramètres d'entrée).

Application à CATHARE

- On suppose que l'on ne connaît rien du modèle



- Pour la stratification contrôlée, il faut construire une surface de réponse avec le moins de calculs possible (problème : 53 paramètres d'entrée).



Phase préliminaire de « screening » (analyse de sensibilité grossière) :
parmi k facteurs, identifier les h facteurs influents ($h \ll k$).

- Les techniques les moins coûteuses (bifurcations séquentielles, group screening) nécessitent de connaître le signe des effets des facteurs.

Application à CATHARE

- On suppose que l'on ne connaît rien du modèle

 Pour la stratification contrôlée, il faut construire une surface de réponse avec le moins de calculs possible (problème : 53 paramètres d'entrée).

 Phase préliminaire de « screening » (analyse de sensibilité grossière) :
parmi k facteurs, identifier les h facteurs influents ($h \ll k$).

- Les techniques les moins coûteuses (bifurcations séquentielles, group screening) nécessitent de connaître le signe des effets des facteurs.
- **Technique des matrices supersaturées :**
 - k (nb de facteurs) $\gg N$ (nb d'expériences) : qq effets prédominants
 - Les interactions entre facteurs ont des effets faibles.

On se restreint à l'utilisation de plans à 2 niveaux.

- Les effets des facteurs sont monotones.

Pour les niveaux bas et haut, on prend les quantiles à 25% et 75%.

Matrice supersaturée

Détails de l'algorithme de construction : Lin (1991), superRoger



- On construit une **matrice d'Hadamard (Plackett-Burman)** (plan qui permet de déterminer les poids de chaque facteur)

($2^{59} // 60$) ; 60 expériences

- On prend une colonne de branchement (59ème)
- On garde toutes les expériences où la colonne 59 est au niveau bas

➔ **Matrice supersaturée (= moitié de la matrice d'Hadamard).**
($2^{53} // 30$) ; 30 expériences

- Analyse de la matrice supersaturée : stepwise regression, algorithmes génétiques, toutes les régressions, ...

Calculs CATHARE

- Résultats du plan saturé (ss) :



<i>N°</i>	<i>Phénomène</i>	<i>Paramètre physique</i>	<i>effet</i>
19	Fuel thermal behaviour	UO ₂ conductivity (fuel rod)	-53
44	Heat transfer in dry zone	Film boiling HTC	-48
9	Fuel thermal behaviour	Axial & radial peaking factors (rod)	32
42	DNB	Critical heat flux (core)	-31
20	Fuel thermal behaviour	UO ₂ specific heat (fuel rod)	28

Calculs CATHARE

- Résultats du plan supersaturé (ss) :



<i>N°</i>	<i>Phénomène</i>	<i>Paramètre physique</i>	<i>effet</i>
19	Fuel thermal behaviour	UO ₂ conductivity (fuel rod)	-53
44	Heat transfer in dry zone	Film boiling HTC	-48
9	Fuel thermal behaviour	Axial & radial peaking factors (rod)	32
42	DNB	Critical heat flux (core)	-31
20	Fuel thermal behaviour	UO ₂ specific heat (fuel rod)	28

- Pour construire une surface de réponse, il faut un plan spécifique :
Par ex. un plan / 5 facteurs influents : un plan composite pour obtenir un polynôme du 2nd degré (+ 27 expériences).

Calculs CATHARE

- Résultats du plan saturé (ss) :



<i>N°</i>	<i>Phénomène</i>	<i>Paramètre physique</i>	<i>effet</i>
19	Fuel thermal behaviour	UO ₂ conductivity (fuel rod)	-53
44	Heat transfer in dry zone	Film boiling HTC	-48
9	Fuel thermal behaviour	Axial & radial peaking factors (rod)	32
42	DNB	Critical heat flux (core)	-31
20	Fuel thermal behaviour	UO ₂ specific heat (fuel rod)	28

- Pour construire une surface de réponse, il faut un plan spécifique :
Par ex. un plan / 5 facteurs influents : un plan composite pour obtenir un polynôme du 2nd degré (+ 27 expériences).
- Pour ne pas refaire de calculs, on choisit une stratégie différente :
 - On ajuste un modèle linéaire à 5 paramètres sur la base du plan ss ;
 - On ajoute progressivement d'autres facteurs dans le modèle de régression, on les garde ou on les jette par analyse des t-values.

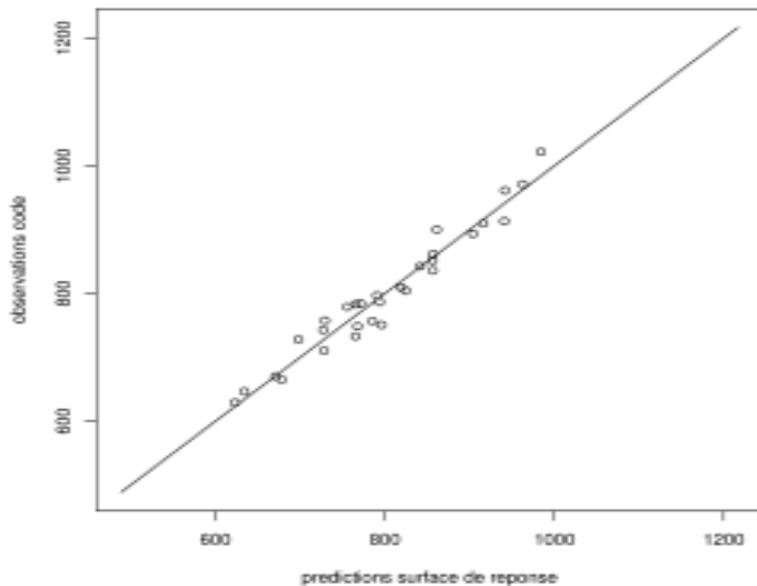
Analyse de la validité de la surface de réponse

Conclusion : méthode fruste mais qui donne des résultats (en 30 calculs).

$$f_r(X) = 660 - 62X_2 + 6X_6 + 590X_9 + 81X_{11} - 405X_{19} \\ + 264X_{20} + 27X_{35} + 6X_{37} - 256X_{42} - 32X_{44}$$

Validation en prédiction (par bootstrap sur la base supersaturée) :

$$R^2 = 0.77 \quad ; \quad \text{IC à 90\% : [0.46, 0.92]}$$



Base des 30 calculs (plan supersaturé)

$$R^2 = 0.95$$

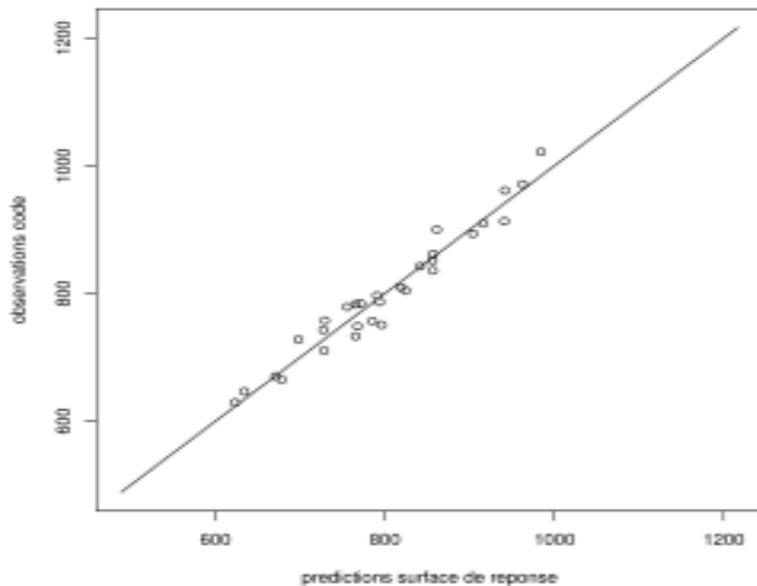
, estimation de quantiles - 10/06

Analyse de la validité de la surface de réponse

Conclusion : méthode fruste mais qui donne des résultats (en 30 calculs).

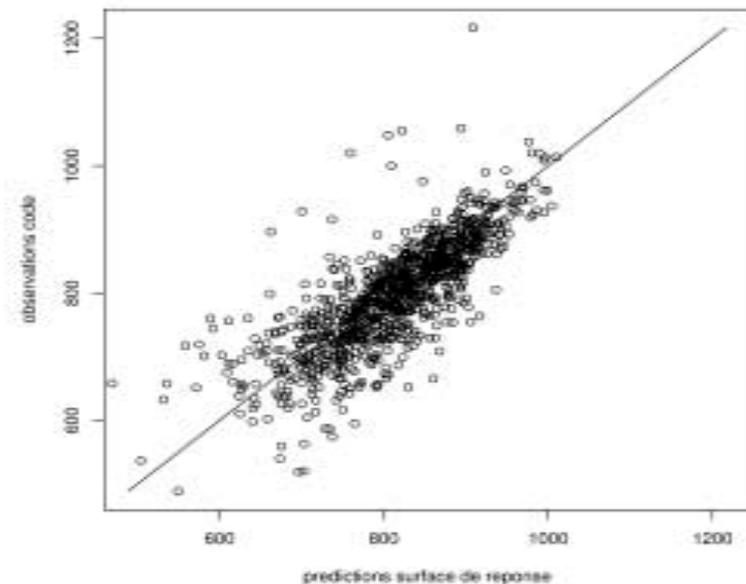
$$f_r(X) = 660 - 62X_2 + 6X_6 + 590X_9 + 81X_{11} - 405X_{19} \\ + 264X_{20} + 27X_{35} + 6X_{37} - 256X_{42} - 32X_{44}$$

Validation en prédiction (par bootstrap sur la base supersaturée) :
 $R^2 = 0.77$; IC à 90% : [0.46, 0.92]



Base des 30 calculs (plan supersaturé)

$$R^2 = 0.95$$



Base supplémentaire (1000 calculs)

$$\rho = 0.66$$

, esti

Stratification contrôlée sur CATHARE

$\alpha = 0.95$; $n = 200$; stratégie à 4 strates.



Strates déterminées à l'aide de 10^7 calculs sur $Z = f_r(X)$:

$$\alpha = 0.95 \longrightarrow z_\alpha = 936.05^\circ\text{C}$$

$$\alpha = 0.85 \longrightarrow z_\alpha = 890.30^\circ\text{C}$$

$$\alpha = 0.55 \longrightarrow z_\alpha = 821.15^\circ\text{C}$$

Stratification contrôlée sur CATHARE

$\alpha = 0.95$; $n=200$; stratégie à 4 strates.



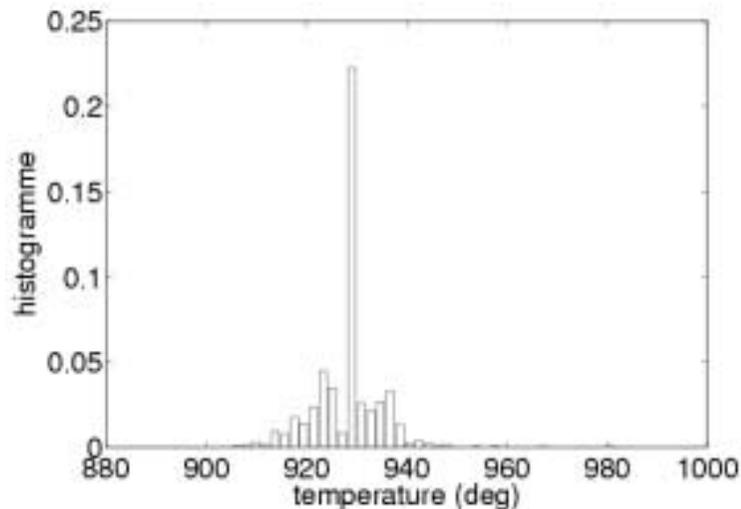
Strates déterminées à l'aide de 10^7 calculs sur $Z=f_r(X)$:

$$\alpha = 0.95 \longrightarrow z_{\alpha} = 936.05^{\circ}\text{C}$$

$$\alpha = 0.85 \longrightarrow z_{\alpha} = 890.30^{\circ}\text{C}$$

$$\alpha = 0.55 \longrightarrow z_{\alpha} = 821.15^{\circ}\text{C}$$

essai 1



Estimation du quantile : 928°C

Écart-type : 7°C

Stratification contrôlée sur CATHARE

$\alpha = 0.95$; $n = 200$; stratégie à 4 strates.



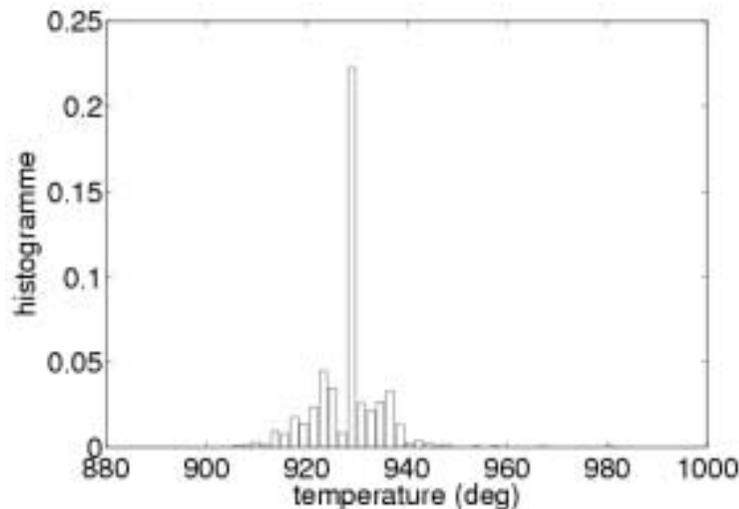
Strates déterminées à l'aide de 10^7 calculs sur $Z = f_r(X)$:

$$\alpha = 0.95 \longrightarrow z_\alpha = 936.05^\circ\text{C}$$

$$\alpha = 0.85 \longrightarrow z_\alpha = 890.30^\circ\text{C}$$

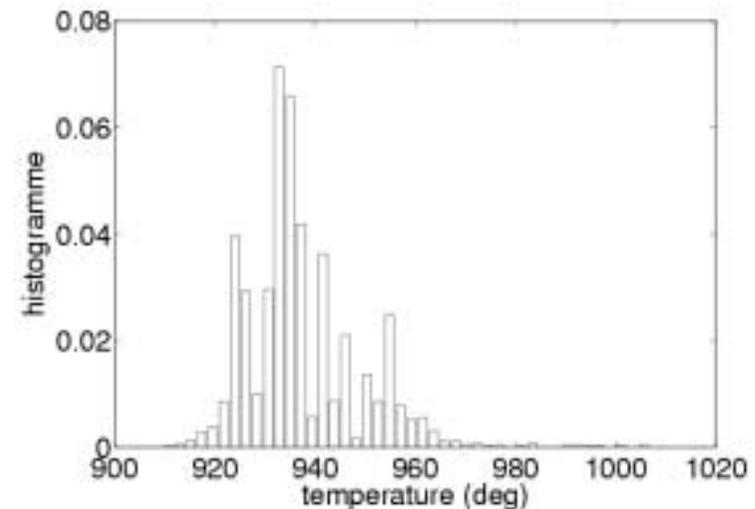
$$\alpha = 0.55 \longrightarrow z_\alpha = 821.15^\circ\text{C}$$

essai 1



Estimation du quantile : 928°C
Écart-type : 7°C

essai 2



Estimation du quantile : 934°C
Écart-type : 11°C

Analyse des résultats

- En regroupant les 2 tests, on a $n=400$ et on obtient :



Estimation du quantile : 930°C
Écart-type : 5.5°C

Analyse des résultats

- En regroupant les 2 tests, on a $n=400$ et on obtient :



Estimation du quantile : 930°C
Écart-type : 5.5°C

- On réalise 1000 calculs supplémentaires.
On obtient pour différentes tailles de l'échantillon (écart-type obtenu avec des répliques générées à partir de la base des 1000 calculs) :

n	100	200	400	600	800	1000
Moyenne estimations	924	926	927	928	928	928
Écart-type estimations	16.7	12.3	9.2	7.6	6.8	6.3

Écart-type obtenue avec la stratification contrôlée :

	11	7	5.5
	(essai 2, n=200)	(essai 1, n=200)	(n=400)

Conclusions / Perspectives

- **La méthode par stratification contrôlée est prometteuse.** Elle utilise toute la spécificité du problème (corrélation code/SR, coût(SR) ~ 0).



- En pratique, robustesse en fonction de n et de α ?
- Choix du nb de strates (en fct de n et de la qualité de la SR) ?
- Comparer estimateur de Wilks (niveau de confiance sur le quantile) avec estimateur par stratification contrôlée avec bootstrap.
- ...

Conclusions / Perspectives

- **La méthode par stratification contrôlée est prometteuse.** Elle utilise toute la spécificité du problème (corrélation code/SR, coût(SR) ~ 0).



- En pratique, robustesse en fonction de n et de α ?
 - Choix du nb de strates (en fct de n et de la qualité de la SR) ?
 - Comparer estimateur de Wilks (niveau de confiance sur le quantile) avec estimateur par stratification contrôlée avec bootstrap.
 - ...
- Echantillonnage préférentiel.

Conclusions / Perspectives

- La méthode par stratification contrôlée est prometteuse. Elle utilise toute la spécificité du problème (corrélation code/SR, coût(SR) ~ 0).



- En pratique, robustesse en fonction de n et de α ?
 - Choix du nb de strates (en fct de n et de la qualité de la SR) ?
 - Comparer estimateur de Wilks (niveau de confiance sur le quantile) avec estimateur par stratification contrôlée avec bootstrap.
 - ...
- Echantillonnage préférentiel.
 - Stratégie optimale (coût minimal) de construction d'une surface de réponse assez grossière (e.g. $R^2 = 0.8$) ?
Pour CATHARE, on pourrait regarder des modèles + riches (e.g. krigeage).

Conclusions / Perspectives

- **La méthode par stratification contrôlée est prometteuse.** Elle utilise toute la spécificité du problème (corrélation code/SR, coût(SR) ~ 0).



- En pratique, robustesse en fonction de n et de α ?
 - Choix du nb de strates (en fct de n et de la qualité de la SR) ?
 - Comparer estimateur de Wilks (niveau de confiance sur le quantile) avec estimateur par stratification contrôlée avec bootstrap.
 - ...
- Echantillonnage préférentiel.
 - Stratégie optimale (coût minimal) de construction d'une surface de réponse assez grossière (e.g. $R^2 = 0.8$) ?
Pour CATHARE, on pourrait regarder des modèles + riches (e.g. krigeage).
 - Plans d'expériences adaptatifs ?