Analyse de sensibilité pour des modèles à entrée corrélées : Application aux modèles de cinétique chimique

Sébastien Da-Veiga

[1]Institut Français du Pétrole, Lyon

[2]Laboratoire de Statistiques et Probabilités
Université Paul Sabatier, Toulouse

Anestis Antoniadis, UJF Grenoble
Fabrice Gamboa, UPS Toulouse
François Wahl, IFP Lyon

-

Planification d'expériences et analyse d'incertitudes pour les gros codes numériques : Approches Stochastiques. 2 et 3 Février 2006

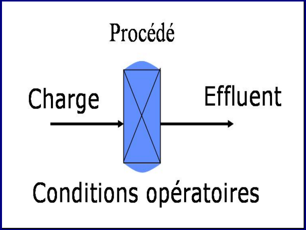
Plan

1. Analyse de sensibilité pour les modèles de cinétique chimique : problématique
2. Estimation des indices de sensibilité par polynômes locaux
3. Résultats sur l'isomérisation du n-butane
4. Conclusions

Plan

Modèles de cinétique chimique

Modèles de cinétique chimique



Modèles de cinétique chimique

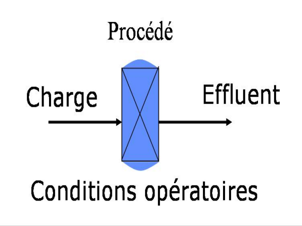


Schéma réactionnel

$$\text{Charge} \xrightarrow[\text{cinétiques}]{\text{Paramètres}} \text{Effluent}$$

Modèles de cinétique chimique

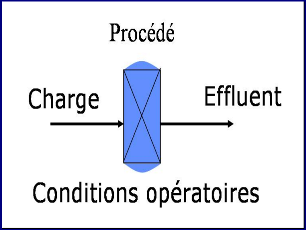


Schéma réactionnel

Charge $\longrightarrow$ Effluent

Paramètres cinétiques

Modèle $Y = f(c, X)$

Y : Composition effluent
c : Conditions opératoires et de charge
X : Paramètres cinétiques (facteurs préexponentiels, énergies d'activation, ...)
f : fonction modélisant la réaction (code de calcul)

Modèles de cinétique chimique

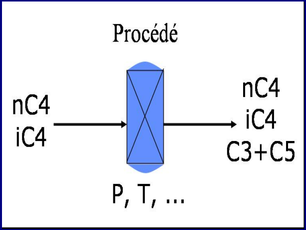


Schéma réactionnel

Charge $\longrightarrow$ Effluent

Paramètres cinétiques

Modèle $Y = f(c, X)$

Y : Composition effluent
c : Conditions opératoires et de charge
X : Paramètres cinétiques (facteurs préexponentiels, énergies d'activation, . . .)
f : fonction modélisant la réaction (code de calcul)

Modèles de cinétique chimique

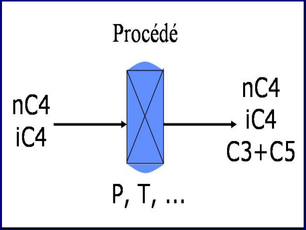


Schéma réactionnel

$$\begin{array}{ccc} nC4 & \longleftrightarrow & iC4 \\ 2\ iC4 & \longrightarrow & C3+C5 \\ nC4+iC4 & \longrightarrow & C3+C5 \\ & (X_1,\ldots,X_8) & \end{array}$$

Modèle $Y = f(c, X)$

Y : Composition effluent
c : Conditions opératoires et de charge
X : Paramètres cinétiques (facteurs préexponentiels, énergies d'activation, ...)
f : fonction modélisant la réaction (code de calcul)

Modèles de cinétique chimique

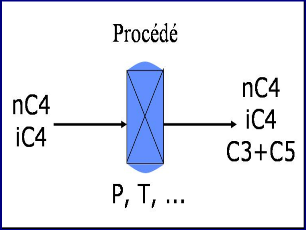


Schéma réactionnel

$$
\begin{array}{ccc}
\text{nC4} & \longleftrightarrow & \text{iC4} \\
\text{2 iC4} & \longrightarrow & \text{C3+C5} \\
\text{nC4+iC4} & \longrightarrow & \text{C3+C5} \\
 & (X_1, \ldots, X_8) &
\end{array}
$$

Modèle $Y = f(c, X)$

Y :[nC4, iC4, C3+C5]
c :[nC4, iC4, P, T, . . .]
X : Paramètres cinétiques (facteurs préexponentiels, énergies d'activation, . . .)
f : fonction modélisant la réaction (code de calcul)

Modèles de cinétique chimique

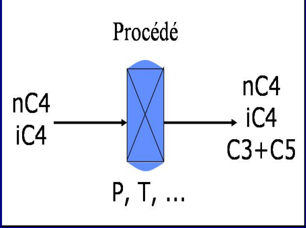


Schéma réactionnel

$$\begin{array}{ccc} \text{nC4} & \longleftrightarrow & \text{iC4} \\ \text{2 iC4} & \longrightarrow & \text{C3+C5} \\ \text{nC4+iC4} & \longrightarrow & \text{C3+C5} \\ & (X_1, \ldots, X_8) & \end{array}$$

Modèle $Y = f(c, X)$

Y :[nC4, iC4, C3+C5]

c :[nC4, iC4, P, T, . . .]

X :$(X_1, X_2, \ldots, X_8)$ Estimés à partir de mesures expérimentales (moindres carrés)

f : fonction modélisant la réaction (code de calcul)

Outils classiques en analyse de sensibilité

$$Y = f(X_1, X_2, \dots, X_d)$$

- Analyse de sensibilité globale : méthodes de variance.
  - Indices de sensibilité du premier ordre :

$$S_i = \frac{\mathrm{Var}(\mathbb{E}(Y|X_i))}{\mathrm{Var}(Y)}$$

  - Indices de sensibilité du deuxième ordre :

$$S_{ij} = \frac{\mathrm{Var}(\mathbb{E}(Y|X_i, X_j)) - \mathrm{Var}(\mathbb{E}(Y|X_i)) - \mathrm{Var}(\mathbb{E}(Y|X_j))}{\mathrm{Var}(Y)}$$

  - $\dots$
- Estimation par méthode de Sobol ou Fast
- Hypothèse : les paramètres d'entrée $(X_1, \dots, X_d)$ sont indépendants

Outils classiques en analyse de sensibilité

$$Y = f(X_1, X_2, \ldots, X_d)$$

- Analyse de sensibilité globale : méthodes de variance.
  - Indices de sensibilité du premier ordre :

$$S_i = \frac{\mathrm{Var}(\mathbb{E}(Y|X_i))}{\mathrm{Var}(Y)}$$

  - Indices de sensibilité du deuxième ordre :

$$S_{ij} = \frac{\mathrm{Var}(\mathbb{E}(Y|X_i, X_j)) - \mathrm{Var}(\mathbb{E}(Y|X_i)) - \mathrm{Var}(\mathbb{E}(Y|X_j))}{\mathrm{Var}(Y)}$$

  - ...
- Estimation par méthode de Sobol ou Fast
- Hypothèse : les paramètres d'entrée $(X_1, \ldots, X_d)$ sont indépendants

Outils classiques en analyse de sensibilité

$$Y = f(X_1, X_2, \ldots, X_d)$$

- Analyse de sensibilité globale : méthodes de variance.
  - Indices de sensibilité du premier ordre :

$$S_i = \frac{\mathrm{Var}(\mathbb{E}(Y|X_i))}{\mathrm{Var}(Y)}$$

  - Indices de sensibilité du deuxième ordre :

$$S_{ij} = \frac{\mathrm{Var}(\mathbb{E}(Y|X_i, X_j)) - \mathrm{Var}(\mathbb{E}(Y|X_i)) - \mathrm{Var}(\mathbb{E}(Y|X_j))}{\mathrm{Var}(Y)}$$

  - ...
- Estimation par méthode de Sobol ou Fast
- Hypothèse : les paramètres d'entrée $(X_1, \ldots, X_d)$ sont indépendants

Outils classiques en analyse de sensibilité

$$Y = f(X_1, X_2, \ldots, X_d)$$

- Analyse de sensibilité globale : méthodes de variance.
  - Indices de sensibilité du premier ordre :

$$S_i = \frac{\mathrm{Var}(\mathbb{E}(Y|X_i))}{\mathrm{Var}(Y)}$$

  - Indices de sensibilité du deuxième ordre :

$$S_{ij} = \frac{\mathrm{Var}(\mathbb{E}(Y|X_i, X_j)) - \mathrm{Var}(\mathbb{E}(Y|X_i)) - \mathrm{Var}(\mathbb{E}(Y|X_j))}{\mathrm{Var}(Y)}$$

  - ...
- Estimation par méthode de Sobol ou Fast
- Hypothèse : les paramètres d'entrée $(X_1, \ldots, X_d)$ sont indépendants

Outils classiques en analyse de sensibilité

$$Y = f(X_1, X_2, \ldots, X_d)$$

- Analyse de sensibilité globale : méthodes de variance.
  - Indices de sensibilité du premier ordre :

$$S_i = \frac{\mathrm{Var}(\mathbb{E}(Y|X_i))}{\mathrm{Var}(Y)}$$

  - Indices de sensibilité du deuxième ordre :

$$S_{ij} = \frac{\mathrm{Var}(\mathbb{E}(Y|X_i, X_j)) - \mathrm{Var}(\mathbb{E}(Y|X_i)) - \mathrm{Var}(\mathbb{E}(Y|X_j))}{\mathrm{Var}(Y)}$$

  - ...
- Estimation par méthode de Sobol ou Fast
- Hypothèse : les paramètres d'entrée $(X_1, \ldots, X_d)$ sont indépendants

Outils classiques en analyse de sensibilité

$$Y = f(X_1, X_2, \ldots, X_d)$$

- Analyse de sensibilité globale : méthodes de variance.
  - Indices de sensibilité du premier ordre :

$$S_i = \frac{\mathrm{Var}(\mathbb{E}(Y|X_i))}{\mathrm{Var}(Y)}$$

  - Indices de sensibilité du deuxième ordre :

$$S_{ij} = \frac{\mathrm{Var}(\mathbb{E}(Y|X_i, X_j)) - \mathrm{Var}(\mathbb{E}(Y|X_i)) - \mathrm{Var}(\mathbb{E}(Y|X_j))}{\mathrm{Var}(Y)}$$

  - ...
- Estimation par méthode de Sobol ou Fast
- Hypothèse : les paramètres d'entrée $(X_1, \ldots, X_d)$ sont indépendants

Indépendance des paramètres en cinétique ?

Rappel

$$X = \underset{\theta \in \mathbb{R}^d}{\operatorname{argmin}} \sum_{i=1}^{n} \omega_i (Y_i - f(c_i, \theta))^2$$

est un *M-estimateur*.

L'hypothèse de Sobol et Fast n'est pas vérifiée

Indépendance des paramètres en cinétique ?

Rappel

$$X = \underset{\theta \in \mathbb{R}^d}{\operatorname{argmin}} \sum_{i=1}^{n} \omega_i (Y_i - f(c_i, \theta))^2$$

est un *M-estimateur*.

L'hypothèse de Sobol et Fast n'est pas vérifiée

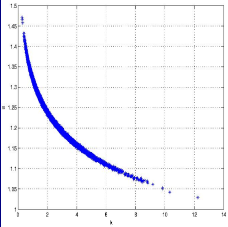
Indépendance des paramètres en cinétique ?

Rappel

$$X = \underset{\theta\in\mathbb{R}^d}{\operatorname{argmin}} \sum_{i=1}^{n} \omega_i (Y_i - f(c_i,\theta))^2$$

est un *M-estimateur*.

Ses composantes sont corrélées



L'hypothèse de Sobol et Fast n'est pas vérifiée

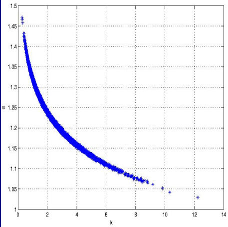
Indépendance des paramètres en cinétique ?

Rappel

$$X = \underset{\theta \in \mathbb{R}^d}{\operatorname{argmin}} \sum_{i=1}^{n} \omega_i (Y_i - f(c_i, \theta))^2$$

est un *M-estimateur*.

Ses composantes sont corrélées



L'hypothèse de Sobol et Fast n'est pas vérifiée

Le cas des paramètres corrélés

- Estimation proposée par Ratto, .., ... mais qui suppose connues les densités conditionnelles $Y|X_i$

Ce n'est pas le cas dans les modèles cinétiques étudiés

- Définition d'indices de sensibilité multidimensionnels (Jacques) :
  - On groupe les variables d'entrées corrélées entre elles *ex :* $((X_1, X_4), (X_2, X_5, X_6), (X_3))$
  - On calcule l'indice de sensibilité de chaque groupe *ex :* S_{14}, S_{256}, S_3
  - On analyse les résultats groupes par groupes

Dans les modèles cinétiques tous les paramètres sont corrélés entre eux, on ne peut isoler des groupes de paramètres indépendants

- Il faut donc trouver une nouvelle méthode d'estimation

Le cas des paramètres corrélés

- Estimation proposée par Ratto, .., ... mais qui suppose connues les densités conditionnelles $Y|X_i$

Ce n'est pas le cas dans les modèles cinétiques étudiés

- Définition d'indices de sensibilité multidimensionnels (Jacques) :
  - On groupe les variables d'entrées corrélées entre elles
    ex : $((X_1, X_4), (X_2, X_5, X_6), (X_3))$
  - On calcule l'indice de sensibilité de chaque groupe
    ex : S_{14}, S_{256}, S_3
  - On analyse les résultats groupes par groupes

Dans les modèles cinétiques tous les paramètres sont corrélés entre eux, on ne peut isoler des groupes de paramètres indépendants

- Il faut donc trouver une nouvelle méthode d'estimation

Le cas des paramètres corrélés

- Estimation proposée par Ratto, .., ... mais qui suppose connues les densités conditionnelles $Y|X_i$

Ce n'est pas le cas dans les modèles cinétiques étudiés

- Définition d'indices de sensibilité multidimensionnels (Jacques) :
  - On groupe les variables d'entrées corrélées entre elles *ex :* $((X_1, X_4), (X_2, X_5, X_6), (X_3))$
  - On calcule l'indice de sensibilité de chaque groupe *ex :* S_{14}, S_{256}, S_3
  - On analyse les résultats groupes par groupes

Dans les modèles cinétiques tous les paramètres sont corrélés entre eux, on ne peut isoler des groupes de paramètres indépendants

- Il faut donc trouver une nouvelle méthode d'estimation

Le cas des paramètres corrélés

- Estimation proposée par Ratto, .., ... mais qui suppose connues les densités conditionnelles $Y|X_i$

Ce n'est pas le cas dans les modèles cinétiques étudiés

- Définition d'indices de sensibilité multidimensionnels (Jacques) :
  - On groupe les variables d'entrées corrélées entre elles *ex :* $((X_1, X_4), (X_2, X_5, X_6), (X_3))$
  - On calcule l'indice de sensibilité de chaque groupe *ex :* S_{14}, S_{256}, S_3
  - On analyse les résultats groupes par groupes

Dans les modèles cinétiques tous les paramètres sont corrélés entre eux, on ne peut isoler des groupes de paramètres indépendants

- Il faut donc trouver une nouvelle méthode d'estimation

Le cas des paramètres corrélés

- Estimation proposée par Ratto, .., ... mais qui suppose connues les densités conditionnelles $Y|X_i$

Ce n'est pas le cas dans les modèles cinétiques étudiés

- Définition d'indices de sensibilité multidimensionnels (Jacques) :
  - On groupe les variables d'entrées corrélées entre elles
    ex : $((X_1, X_4), (X_2, X_5, X_6), (X_3))$
  - On calcule l'indice de sensibilité de chaque groupe
    ex : S_{14}, S_{256}, S_3
  - On analyse les résultats groupes par groupes

Dans les modèles cinétiques tous les paramètres sont corrélés entre eux, on ne peut isoler des groupes de paramètres indépendants

- Il faut donc trouver une nouvelle méthode d'estimation

Le cas des paramètres corrélés

- Estimation proposée par Ratto, .., ... mais qui suppose connues les densités conditionnelles $Y|X_i$

Ce n'est pas le cas dans les modèles cinétiques étudiés

- Définition d'indices de sensibilité multidimensionnels (Jacques) :
  - On groupe les variables d'entrées corrélées entre elles
    ex : $((X_1, X_4), (X_2, X_5, X_6), (X_3))$
  - On calcule l'indice de sensibilité de chaque groupe
    ex : S_{14}, S_{256}, S_3
  - On analyse les résultats groupes par groupes

Dans les modèles cinétiques tous les paramètres sont corrélés entre eux, on ne peut isoler des groupes de paramètres indépendants

- Il faut donc trouver une nouvelle méthode d'estimation

Le cas des paramètres corrélés

- Estimation proposée par Ratto, .., ... mais qui suppose connues les densités conditionnelles $Y|X_i$

Ce n'est pas le cas dans les modèles cinétiques étudiés

- Définition d'indices de sensibilité multidimensionnels (Jacques) :
  - On groupe les variables d'entrées corrélées entre elles
    ex : $((X_1, X_4), (X_2, X_5, X_6), (X_3))$
  - On calcule l'indice de sensibilité de chaque groupe
    ex : S_{14}, S_{256}, S_3
  - On analyse les résultats groupes par groupes

Dans les modèles cinétiques tous les paramètres sont corrélés entre eux, on ne peut isoler des groupes de paramètres indépendants

- Il faut donc trouver une nouvelle méthode d'estimation

Le cas des paramètres corrélés

- Estimation proposée par Ratto, .., ... mais qui suppose connues les densités conditionnelles $Y|X_i$

Ce n'est pas le cas dans les modèles cinétiques étudiés

- Définition d'indices de sensibilité multidimensionnels (Jacques) :
  - On groupe les variables d'entrées corrélées entre elles *ex :* $((X_1, X_4), (X_2, X_5, X_6), (X_3))$
  - On calcule l'indice de sensibilité de chaque groupe *ex :* S_{14}, S_{256}, S_3
  - On analyse les résultats groupes par groupes

Dans les modèles cinétiques tous les paramètres sont corrélés entre eux, on ne peut isoler des groupes de paramètres indépendants

- Il faut donc trouver une nouvelle méthode d'estimation

Le cas des paramètres corrélés

- Estimation proposée par Ratto, .., ... mais qui suppose connues les densités conditionnelles $Y|X_i$

Ce n'est pas le cas dans les modèles cinétiques étudiés

- Définition d'indices de sensibilité multidimensionnels (Jacques) :
  - On groupe les variables d'entrées corrélées entre elles *ex :* $((X_1, X_4), (X_2, X_5, X_6), (X_3))$
  - On calcule l'indice de sensibilité de chaque groupe *ex :* S_{14}, S_{256}, S_3
  - On analyse les résultats groupes par groupes

Dans les modèles cinétiques tous les paramètres sont corrélés entre eux, on ne peut isoler des groupes de paramètres indépendants

- Il faut donc trouver une nouvelle méthode d'estimation

Plan

1. Analyse de sensibilité pour les modèles de cinétique chimique : problématique
2. Estimation des indices de sensibilité par polynômes locaux
3. Résultats sur l'isomérisation du n-butane
4. Conclusions

Formulation du problème

- On cherche à estimer une quantité de la forme

$$\frac{\mathrm{Var}(\mathbb{E}(Y|X))}{\mathrm{Var}(Y)}$$

  où (X, Y) est un vecteur aléatoire dont on a un échantillon $(X_i, Y_i)_{i=1,\dots,n}$.
- On va commencer par estimer l'espérance conditionnelle $\mathbb{E}(Y|X)$ à partir de l'échantillon $(X_i, Y_i)_{i=1,\dots,n}$
- On propose d'utiliser les polynômes locaux

Formulation du problème

- On cherche à estimer une quantité de la forme

$$\frac{\mathrm{Var}(\mathbb{E}(Y|X))}{\mathrm{Var}(Y)}$$

  où (X, Y) est un vecteur aléatoire dont on a un échantillon $(X_i, Y_i)_{i=1,\ldots,n}$.
- On va commencer par estimer l'espérance conditionnelle $\mathbb{E}(Y|X)$ à partir de l'échantillon $(X_i, Y_i)_{i=1,\ldots,n}$
- On propose d'utiliser les polynômes locaux

Formulation du problème

- On cherche à estimer une quantité de la forme

$$\frac{\mathrm{Var}(\mathbb{E}(Y|X))}{\mathrm{Var}(Y)}$$

  où (X, Y) est un vecteur aléatoire dont on a un échantillon $(X_i, Y_i)_{i=1,\dots,n}$.
- On va commencer par estimer l'espérance conditionnelle $\mathbb{E}(Y|X)$ à partir de l'échantillon $(X_i, Y_i)_{i=1,\dots,n}$
- On propose d'utiliser les polynômes locaux

Estimation de $\mathbb{E}(Y|X)$ (1)

- On écrit un modèle de régression classique de Y sur X

$$Y_i = m(X_i) + \sigma(X_i)\epsilon_i \quad i = 1, \ldots, n$$

- Estimation de la fonction $m(\cdot)$
  - Régression paramétrique
    ex : $m(x) = a + bx$, $m(x) = a\exp(b + cx^2)$, ...
  - Régression non-paramétrique : pas d'hypothèse d'appartenance à une famille de fonctions
    ex : Splines, Polynômes locaux (estimateurs à noyaux), ...

Estimation de $\mathbb{E}(Y|X)$ (1)

- On écrit un modèle de régression classique de Y sur X

$$Y_i = m(X_i) + \sigma(X_i)\epsilon_i \quad i = 1, \ldots, n$$

où $m(x) = \mathbb{E}(Y|X = x)$ et $\sigma^2(x) = \mathrm{Var}(Y|X = x)$.

- Estimation de la fonction $m(\cdot)$
  - Régression paramétrique
    ex : $m(x) = a + bx$, $m(x) = a\exp(b + cx^2)$, ...
  - Régression non-paramétrique : pas d'hypothèse d'appartenance à une famille de fonctions
    ex : Splines, Polynômes locaux (estimateurs à noyaux), ...

Estimation de $\mathbb{E}(Y|X)$ (1)

- On écrit un modèle de régression classique de Y sur X

$$Y_i = m(X_i) + \sigma(X_i)\epsilon_i \quad i = 1, \ldots, n$$

où $m(x) = \mathbb{E}(Y|X = x)$ et $\sigma^2(x) = \mathrm{Var}(Y|X = x)$.
$(\epsilon_1, \ldots, \epsilon_n)$ erreurs centrées, réduites, indépendantes des X_i.

- Estimation de la fonction $m(\cdot)$
  - Régression paramétrique
    ex : $m(x) = a + bx$, $m(x) = a\exp(b + cx^2)$, ...
  - Régression non-paramétrique : pas d'hypothèse d'appartenance à une famille de fonctions
    ex : Splines, Polynômes locaux (estimateurs à noyaux), ...

Estimation de $\mathbb{E}(Y|X)$ (1)

- On écrit un modèle de régression classique de Y sur X

$$Y_i = m(X_i) + \sigma(X_i)\epsilon_i \quad i = 1, \dots, n$$

  où $m(x) = \mathbb{E}(Y|X = x)$ et $\sigma^2(x) = \mathrm{Var}(Y|X = x)$.
  $(\epsilon_1, \dots, \epsilon_n)$ erreurs centrées, réduites, indépendantes des X_i.

- Estimation de la fonction $m(\cdot)$
  - Régression paramétrique
    ex : $m(x) = a + bx$, $m(x) = a\exp(b + cx^2)$, $\dots$

Si famille paramétrique trop éloignée, mauvaise estimation

  - Régression non-paramétrique : pas d'hypothèse d'appartenance à une famille de fonctions
    ex : Splines, Polynômes locaux (estimateurs à noyaux), $\dots$

Estimation de $\mathbb{E}(Y|X)$ (1)

- On écrit un modèle de régression classique de Y sur X

$$Y_i = m(X_i) + \sigma(X_i)\epsilon_i \quad i = 1, \ldots, n$$

où $m(x) = \mathbb{E}(Y|X = x)$ et $\sigma^2(x) = \mathrm{Var}(Y|X = x)$.
$(\epsilon_1, \ldots, \epsilon_n)$ erreurs centrées, réduites, indépendantes des X_i.

- Estimation de la fonction $m(\cdot)$
  - Régression paramétrique
    ex : $m(x) = a + bx$, $m(x) = a\exp(b + cx^2)$, ...

Si famille paramétrique trop éloignée, mauvaise estimation

  - Régression non-paramétrique : pas d'hypothèse d'appartenance à une famille de fonctions
    ex : Splines, Polynômes locaux (estimateurs à noyaux), ...

Estimation de $\mathbb{E}(Y|X)$ (1)

- On écrit un modèle de régression classique de Y sur X

$$Y_i = m(X_i) + \sigma(X_i)\epsilon_i \quad i = 1, \ldots, n$$

où $m(x) = \mathbb{E}(Y|X = x)$ et $\sigma^2(x) = \mathrm{Var}(Y|X = x)$.
$(\epsilon_1, \ldots, \epsilon_n)$ erreurs centrées, réduites, indépendantes des X_i.

- Estimation de la fonction $m(\cdot)$
  - Régression paramétrique
    ex : $m(x) = a + bx$, $m(x) = a\exp(b + cx^2)$, …

Si famille paramétrique trop éloignée, mauvaise estimation

  - Régression non-paramétrique : pas d'hypothèse d'appartenance à une famille de fonctions
    ex : Splines, Polynômes locaux (estimateurs à noyaux), …

Estimation de $\mathbb{E}(Y|X)$ (1)

- On écrit un modèle de régression classique de Y sur X

$$Y_i = m(X_i) + \sigma(X_i)\epsilon_i \quad i = 1, \dots, n$$

  où $m(x) = \mathbb{E}(Y|X = x)$ et $\sigma^2(x) = \mathrm{Var}(Y|X = x)$.
  $(\epsilon_1, \dots, \epsilon_n)$ erreurs centrées, réduites, indépendantes des X_i.

- Estimation de la fonction $m(\cdot)$
  - Régression paramétrique
    ex : $m(x) = a + bx$, $m(x) = a\exp(b + cx^2)$, $\dots$

Si famille paramétrique trop éloignée, mauvaise estimation

  - Régression non-paramétrique : pas d'hypothèse d'appartenance à une famille de fonctions
    ex : Splines, Polynômes locaux (estimateurs à noyaux), $\dots$

Analyse de sensibilité pour des modèles à entrée corrélées

Sébastien Da-Veiga

Estimation de $\mathbb{E}(Y|X)$ (1)

- On écrit un modèle de régression classique de Y sur X

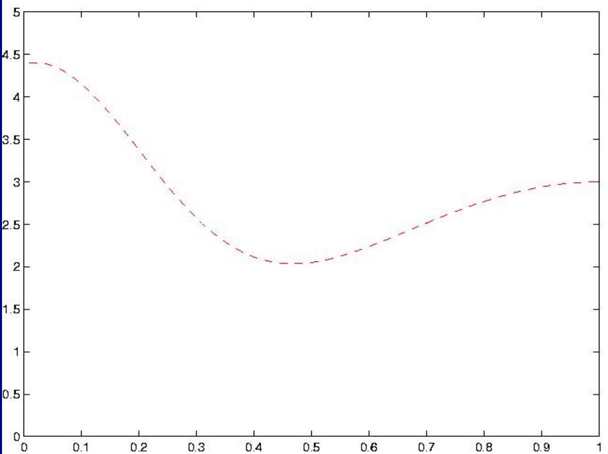
$$Y_i = m(X_i) + \sigma(X_i)\epsilon_i \quad i = 1, \dots, n$$

  où $m(x) = \mathbb{E}(Y|X = x)$ et $\sigma^2(x) = \mathrm{Var}(Y|X = x)$.
  $(\epsilon_1, \dots, \epsilon_n)$ erreurs centrées, réduites, indépendantes des X_i.

- Estimation de la fonction $m(\cdot)$
  - Régression paramétrique
    ex : $m(x) = a + bx$, $m(x) = a\exp(b + cx^2)$, ...

    Si famille paramétrique trop éloignée, mauvaise estimation

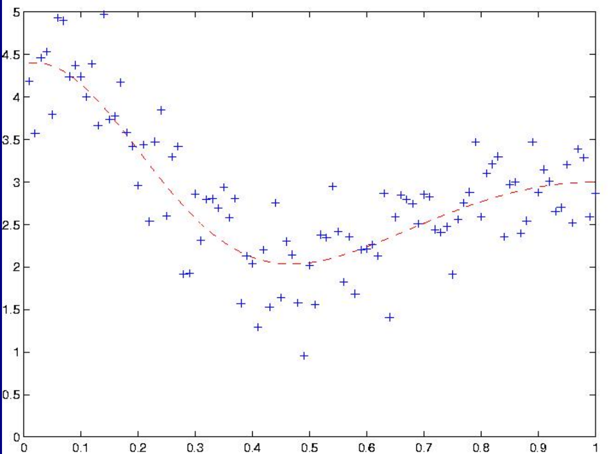
  - Régression non-paramétrique : pas d'hypothèse d'appartenance à une famille de fonctions
    ex : Splines, Polynômes locaux (estimateurs à noyaux), ...



La fonction $m(\cdot)$ est approchée localement par un polynôme d'ordre p

$$m(z) \approx \beta_0 + \beta_1(z-x) + \ldots + \beta_p(z-x)^p$$

pour z dans un voisinage de x



La fonction $m(\cdot)$ est approchée localement par un polynôme d'ordre p

$$m(z) \approx \beta_0 + \beta_1(z-x) + \ldots + \beta_p(z-x)^p$$

pour z dans un voisinage de x

Analyse de sensibilité pour des modèles à entrée corrélées

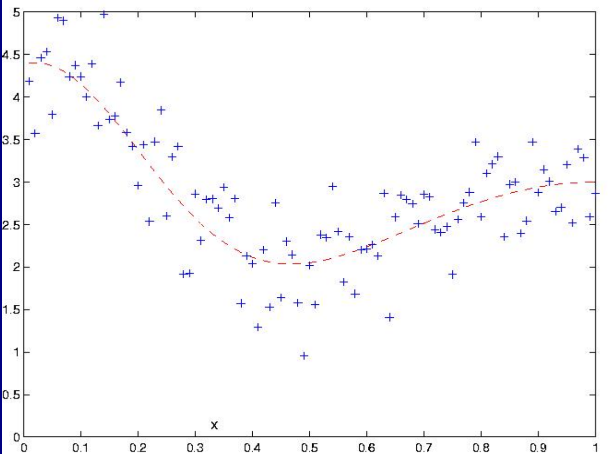
Sébastien DA-VEIGA

Problématique

Polynômes locaux

Application isomérisation

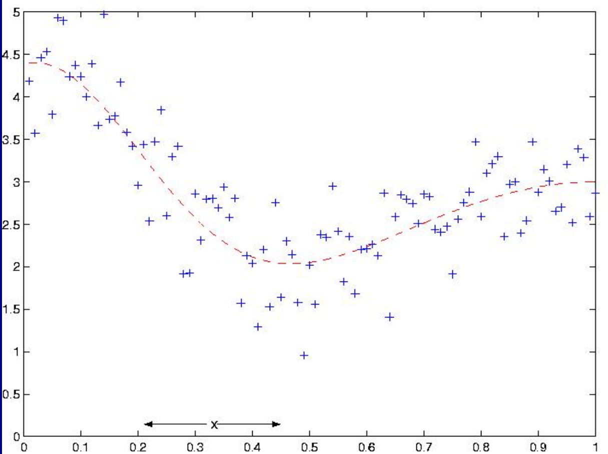
Conclusions



La fonction $m(\cdot)$ est approchée localement par un polynôme d'ordre p

$$m(z) \approx \beta_0 + \beta_1(z-x) + \ldots + \beta_p(z-x)^p$$

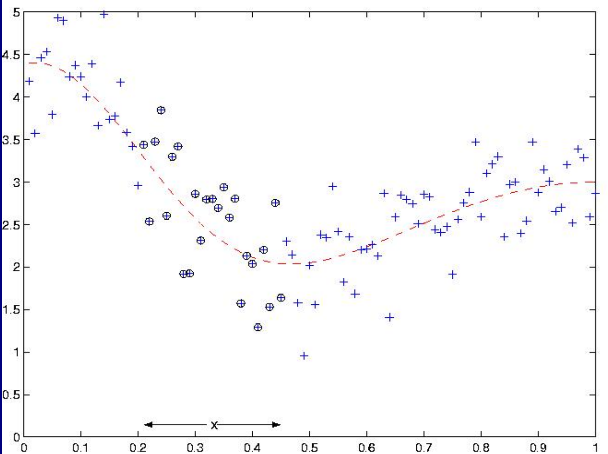
pour z dans un voisinage de x



La fonction $m(\cdot)$ est approchée localement par un polynôme d'ordre p

$$m(z) \approx \beta_0 + \beta_1(z-x) + \ldots + \beta_p(z-x)^p$$

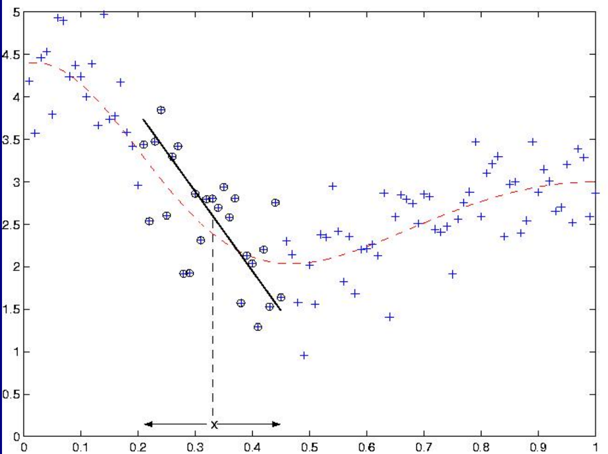
pour z dans un voisinage de x



La fonction $m(\cdot)$ est approchée localement par un polynôme d'ordre p

$$m(z) \approx \beta_0 + \beta_1(z - x) + \ldots + \beta_p(z - x)^p$$

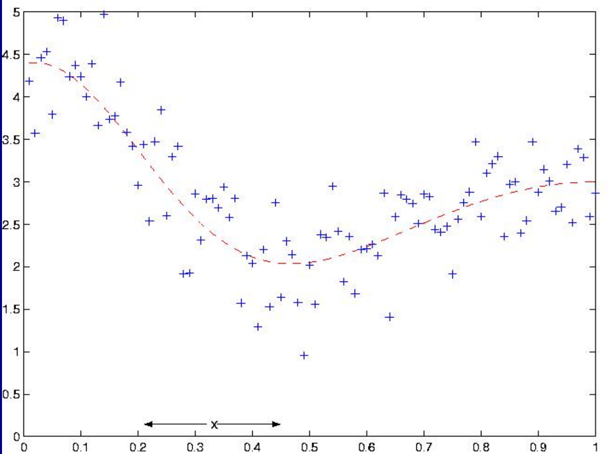
pour z dans un voisinage de x



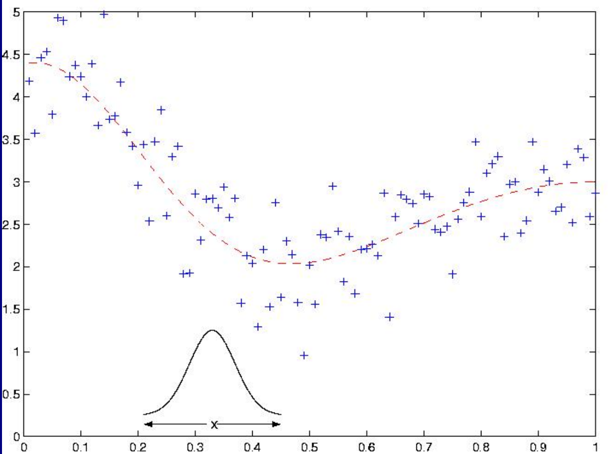
La fonction $m(\cdot)$ est approchée localement par un polynôme d'ordre p

$$m(z) \approx \beta_0 + \beta_1(z-x) + \ldots + \beta_p(z-x)^p$$

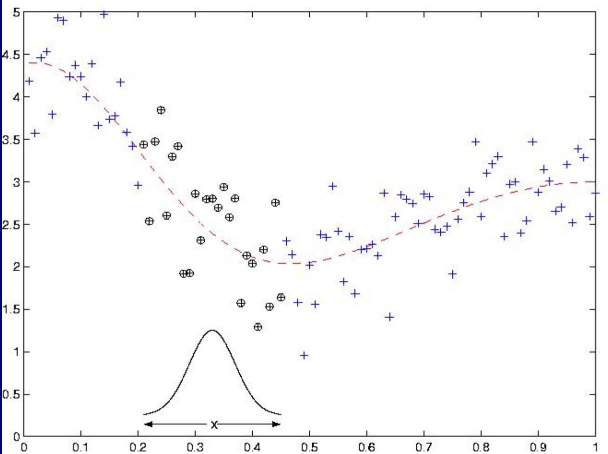
pour z dans un voisinage de x



Définition d'un voisinage : utilisation d'une fonction noyau notée K



Définition d'un voisinage : utilisation d'une fonction noyau notée K



Définition d'un voisinage : utilisation d'une fonction noyau notée K

- Finalement, le polynôme est fitté à partir de l'échantillon $(X_i, Y_i)_{i=1,\dots,n}$ par résolution d'un problème de moindres carrés pondérés

- Trois paramètres à sélectionner :
  - Le noyau K : peu d'influence
  - L'ordre p du polynôme : en pratique 1 ou 2
  - Le paramètre d'échelle h : asymptotique, validation croisée, EBBS, ...

- Finalement, le polynôme est fitté à partir de l'échantillon $(X_i, Y_i)_{i=1,\ldots,n}$ par résolution d'un problème de moindres carrés pondérés

$$\min_{\beta} \sum_{i=1}^{n} \left(Y_i - \sum_{j=0}^{p} \beta_j (X_i - x)^j \right)^2 K\left(\frac{X_i - x}{h} \right)$$

- Trois paramètres à sélectionner :
  - Le noyau K : peu d'influence
  - L'ordre p du polynôme : en pratique 1 ou 2
  - Le paramètre d'échelle h : asymptotique, validation croisée, EBBS, ...

- Finalement, le polynôme est fitté à partir de l'échantillon $(X_i, Y_i)_{i=1,\dots,n}$ par résolution d'un problème de moindres carrés pondérés

$$\min_{\beta} \sum_{i=1}^{n} \left(Y_i - \sum_{j=0}^{p} \beta_j (X_i - x)^{j} \right)^2 K\left(\frac{X_i - x}{h} \right)$$

$$\hat{m}(x) = \hat{\beta}_0(x)$$

- Trois paramètres à sélectionner :
  - Le noyau K : peu d'influence
  - L'ordre p du polynôme : en pratique 1 ou 2
  - Le paramètre d'échelle h : asymptotique, validation croisée, EBBS, ...

- Finalement, le polynôme est fitté à partir de l'échantillon $(X_i, Y_i)_{i=1,\dots,n}$ par résolution d'un problème de moindres carrés pondérés

$$\min_{\beta} \sum_{i=1}^{n} \left(Y_i - \sum_{j=0}^{p} \beta_j (X_i - x)^{\,j} \right)^2 K\left(\frac{X_i - x}{h} \right)$$

$$\hat{m}(x) = \hat{\beta}_0(x)$$

- Trois paramètres à sélectionner :
    - Le noyau K : peu d'influence
    - L'ordre p du polynôme : en pratique 1 ou 2
    - Le paramètre d'échelle h : asymptotique, validation croisée, EBBS, ...

- Finalement, le polynôme est fitté à partir de l'échantillon $(X_i, Y_i)_{i=1,\dots,n}$ par résolution d'un problème de moindres carrés pondérés

$$\min_{\beta} \sum_{i=1}^{n} \left(Y_i - \sum_{j=0}^{p} \beta_j (X_i - x)^j \right)^2 K\left(\frac{X_i - x}{h} \right)$$

$$\hat{m}(x) = \hat{\beta}_0(x)$$

- Trois paramètres à sélectionner :
  - Le noyau K : peu d'influence
  - L'ordre p du polynôme : en pratique 1 ou 2
  - Le paramètre d'échelle h : asymptotique, validation croisée, EBBS, ...

- Finalement, le polynôme est fitté à partir de l'échantillon $(X_i, Y_i)_{i=1,\ldots,n}$ par résolution d'un problème de moindres carrés pondérés

$$\min_{\beta} \sum_{i=1}^{n} \left(Y_i - \sum_{j=0}^{p} \beta_j (X_i - x)^{j} \right)^2 K\left(\frac{X_i - x}{h}\right)$$

$$\hat{m}(x) = \hat{\beta}_0(x)$$

- Trois paramètres à sélectionner :
  - Le noyau K : peu d'influence
  - L'ordre p du polynôme : en pratique 1 ou 2
  - Le paramètre d'échelle h : asymptotique, validation croisée, EBBS, ...

- Finalement, le polynôme est fitté à partir de l'échantillon $(X_i, Y_i)_{i=1,\dots,n}$ par résolution d'un problème de moindres carrés pondérés

$$\min_{\beta} \sum_{i=1}^{n} \left(Y_i - \sum_{j=0}^{p} \beta_j (X_i - x)^j \right)^2 K\left(\frac{X_i - x}{h} \right)$$

$$\hat{m}(x) = \hat{\beta}_0(x)$$

- Trois paramètres à sélectionner :
  - Le noyau K : peu d'influence
  - L'ordre p du polynôme : en pratique 1 ou 2
  - Le paramètre d'échelle h : asymptotique, validation croisée, EBBS, ...

Application aux indices de sensibilité

- On a donc un estimateur $\hat{m}(x)$ de la fonction $m(x) = \mathbb{E}(Y|X = x)$
- On cherche à estimer $\mathrm{Var}(E(Y|X))$
- Si on connaissait la fonction $m(\cdot)$, on pourrait estimer cette quantité avec l'estimateur non biaisé classique de la variance

$$\frac{1}{n'-1}\sum_{j=1}^{n'}\left(m(\tilde{X}_j) - \bar{m}\right)^2$$

où $(\tilde{X}_j)_{j=1,\ldots,n'}$ échantillon i.i.d. de X.

Application aux indices de sensibilité

- On a donc un estimateur $\hat{m}(x)$ de la fonction $m(x) = \mathbb{E}(Y|X = x)$
- On cherche à estimer $\mathrm{Var}(E(Y|X))$
- Si on connaissait la fonction $m(\cdot)$, on pourrait estimer cette quantité avec l'estimateur non biaisé classique de la variance

$$\frac{1}{n'-1}\sum_{j=1}^{n'}\left(m(\tilde{X}_j) - \bar{m}\right)^2$$

où $(\tilde{X}_j)_{j=1,\ldots,n'}$ échantillon i.i.d. de X.

Application aux indices de sensibilité

- On a donc un estimateur $\hat{m}(x)$ de la fonction $m(x) = \mathbb{E}(Y|X = x)$
- On cherche à estimer $\mathrm{Var}(E(Y|X))$
- Si on connaissait la fonction $m(\cdot)$, on pourrait estimer cette quantité avec l'estimateur non biaisé classique de la variance

$$\frac{1}{n' - 1} \sum_{j=1}^{n'} \left(m(\tilde{X}_j) - \bar{m} \right)^2$$

où $(\tilde{X}_j)_{j=1,\dots,n'}$ échantillon i.i.d. de X.

- On ne connaît pas $m(\cdot)$, on va la remplacer par son estimateur dans la formule précédente :

$$\frac{1}{n'-1}\sum_{j=1}^{n'}\left(\hat{m}(\tilde{X}_j)-\hat{\hat{m}}\right)^2$$

- Remarque : on a besoin de deux échantillons
  - Le premier pour estimer $\hat{m}$
  - Le second pour former $(\tilde{X}_j)_{j=1,\dots,n'}$

  Solutions : soit on coupe l'échantillon initial en deux, soit on utilise un échantillon bootstrappé

- On ne connaît pas $m(\cdot)$, on va la remplacer par son estimateur dans la formule précédente :

$$\frac{1}{n'-1}\sum_{j=1}^{n'}\left(\hat{m}(\tilde{X}_j)-\hat{\hat{m}}\right)^2$$

- Remarque : on a besoin de deux échantillons
  - Le premier pour estimer $\hat{m}$
  - Le second pour former $(\tilde{X}_j)_{j=1,\dots,n'}$

Solutions : soit on coupe l'échantillon initial en deux, soit on utilise un échantillon bootstrappé

- On ne connaît pas $m(\cdot)$, on va la remplacer par son estimateur dans la formule précédente :

$$\frac{1}{n'-1}\sum_{j=1}^{n'}\left(\hat{m}(\tilde{X}_j)-\hat{\bar{m}}\right)^2$$

- Remarque : on a besoin de deux échantillons
  - Le premier pour estimer $\hat{m}$
  - Le second pour former $(\tilde{X}_j)_{j=1,\dots,n'}$

Solutions : soit on coupe l'échantillon initial en deux, soit on utilise un échantillon bootstrappé

- On ne connaît pas $m(\cdot)$, on va la remplacer par son estimateur dans la formule précédente :

$$\frac{1}{n'-1}\sum_{j=1}^{n'}\left(\hat{m}(\tilde{X}_j)-\hat{\bar{m}}\right)^2$$

- Remarque : on a besoin de deux échantillons
    - Le premier pour estimer $\hat{m}$
    - Le second pour former $(\tilde{X}_j)_{j=1,\ldots,n'}$

  Solutions : soit on coupe l'échantillon initial en deux, soit on utilise un échantillon bootstrappé

- On ne connaît pas $m(\cdot)$, on va la remplacer par son estimateur dans la formule précédente :

$$\frac{1}{n'-1}\sum_{j=1}^{n'}\left(\hat{m}(\tilde{X}_j)-\hat{\hat{m}}\right)^2$$

- Remarque : on a besoin de deux échantillons
  - Le premier pour estimer $\hat{m}$
  - Le second pour former $(\tilde{X}_j)_{j=1,\dots,n'}$

  Solutions : soit on coupe l'échantillon initial en deux, soit on utilise un échantillon bootstrappé

Résultats théoriques

- On peut montrer que l'estimateur proposé est asympotiquement sans biais
- En utilisant la relation classique $\mathrm{Var}(Y) = \mathbb{E}(\mathrm{Var}(Y|X)) + \mathrm{Var}(\mathbb{E}(Y|X))$ on peut se ramener au calcul de $\mathbb{E}(\mathrm{Var}(Y|X))$.

Résultats théoriques

- On peut montrer que l'estimateur proposé est asympotiquement sans biais
- En utilisant la relation classique $\mathrm{Var}(Y) = \mathbb{E}(\mathrm{Var}(Y|X)) + \mathrm{Var}(\mathbb{E}(Y|X))$ on peut se ramener au calcul de $\mathbb{E}(\mathrm{Var}(Y|X))$.

Résultats théoriques

- On peut montrer que l'estimateur proposé est asympotiquement sans biais
- En utilisant la relation classique $\mathrm{Var}(Y) = \mathbb{E}(\mathrm{Var}(Y|X)) + \mathrm{Var}(\mathbb{E}(Y|X))$ on peut se ramener au calcul de $\mathbb{E}(\mathrm{Var}(Y|X))$.

Résultats théoriques

- On peut montrer que l'estimateur proposé est asympotiquement sans biais
- En utilisant la relation classique $\mathrm{Var}(Y) = \mathbb{E}(\mathrm{Var}(Y|X)) + \mathrm{Var}(\mathbb{E}(Y|X))$ on peut se ramener au calcul de $\mathbb{E}(\mathrm{Var}(Y|X))$.

Résultats théoriques

- On peut montrer que l'estimateur proposé est asympotiquement sans biais
- En utilisant la relation classique $\mathrm{Var}(Y) = \mathbb{E}(\mathrm{Var}(Y|X)) + \mathrm{Var}(\mathbb{E}(Y|X))$ on peut se ramener au calcul de $\mathbb{E}(\mathrm{Var}(Y|X))$.
  Par polynômes locaux on peut estimer $\mathrm{Var}(Y|X)$, puis $\mathbb{E}(\mathrm{Var}(Y|X))$ comme précédemment.

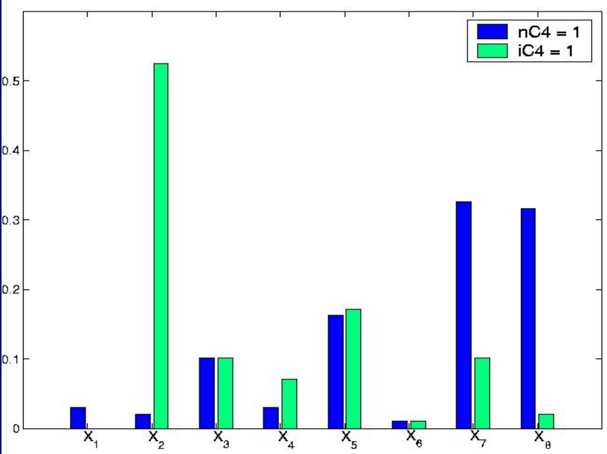
Résultats théoriques

- On peut montrer que l'estimateur proposé est asympotiquement sans biais
- En utilisant la relation classique $\mathrm{Var}(Y) = \mathbb{E}(\mathrm{Var}(Y|X)) + \mathrm{Var}(\mathbb{E}(Y|X))$ on peut se ramener au calcul de $\mathbb{E}(\mathrm{Var}(Y|X))$.
  Par polynômes locaux on peut estimer $\mathrm{Var}(Y|X)$, puis $\mathbb{E}(\mathrm{Var}(Y|X))$ comme précédemment.
  Dans ce cas on obtient un estimateur consistant (asympotiquement sans biais et variance tendant vers 0)

Plan

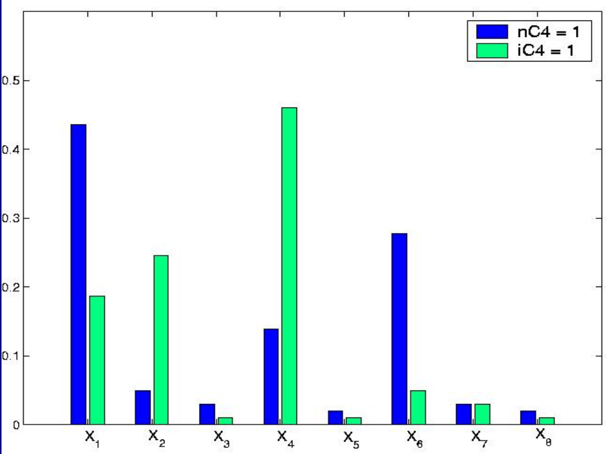
Isomérisation du n-butane

Indices de sensibilité de la sortie C3+C5



Isomérisation du n-butane

Indices de sensibilité de la sortie nC4



Plan

1. Analyse de sensibilité pour les modèles de cinétique chimique : problématique
2. Estimation des indices de sensibilité par polynômes locaux
3. Résultats sur l'isomérisation du n-butane
4. Conclusions

Conclusions

- Estimateurs sans biais des indices de sensibilité du premier ordre pour des paramètres d'entrée corrélés
- Inconvénient : temps de calcul pour le choix du paramètre d'échelle h
- Difficulté d'interprétation dans le cas de paramètres corrélés (sauf si l'un des paramètres est très peu influent)

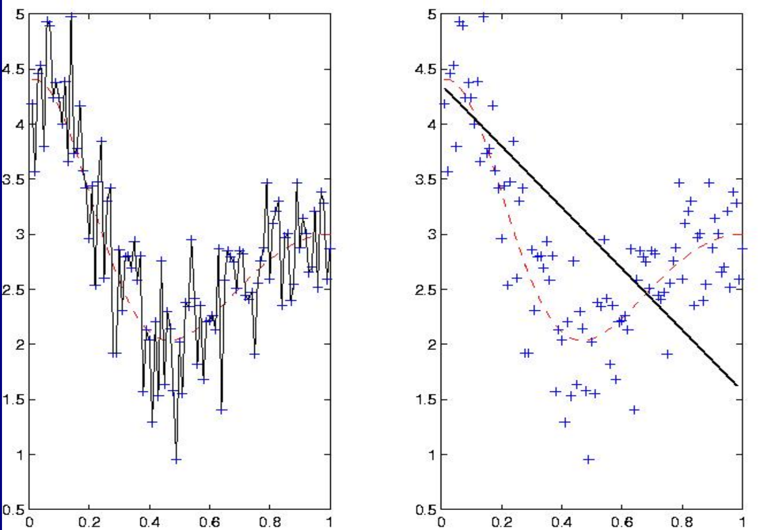
Conclusions

- Estimateurs sans biais des indices de sensibilité du premier ordre pour des paramètres d'entrée corrélés
- Inconvénient : temps de calcul pour le choix du paramètre d'échelle h
- Difficulté d'interprétation dans le cas de paramètres corrélés (sauf si l'un des paramètres est très peu influent)

Conclusions

- Estimateurs sans biais des indices de sensibilité du premier ordre pour des paramètres d'entrée corrélés
- Inconvénient : temps de calcul pour le choix du paramètre d'échelle h
- Difficulté d'interprétation dans le cas de paramètres corrélés (sauf si l'un des paramètres est très peu influent)

Choix de la fenêtre h



Estimation de $\mathrm{Var}(Y|X)$ (1)

- Si $m(\cdot)$ était connue, on pourrait écrire

$$\sigma^2(x) = \mathbb{E}(r^2|X = x)$$

avec $r^2 = (Y - m(X))^2$.

- Donc $\sigma^2(\cdot)$ estimé par polynômes locaux :

$$\min_{\gamma} \sum_{i=1}^{n} \left((Y_i - m(X_i))^2 - \sum_{j=0}^{q} \gamma_j (X_i - x)^j \right)^2 K_2\left(\frac{X_i - x}{h_2}\right)$$

Estimation de $\mathrm{Var}(Y|X)$ (2)

- Mais $m(\cdot)$ inconnue
- On la remplace par son estimateur :

$$\min_{\gamma} \sum_{i=1}^{n} \left((Y_i - \hat{m}(X_i))^2 - \sum_{j=0}^{q} \gamma_j (X_i - x)^j \right)^2 K_2\left(\frac{X_i - x}{h_2}\right)$$

$$\hat{\sigma}^2(x) = \hat{\gamma}_0(x)$$

- Remarques :
  - K_2 et h_2 et q différents de K, h et p
  - Deux estimations successives par polynômes locaux : coût calculatoire

Biais et Variance des estimateurs

- Biais de $\hat{m}(\cdot)$:

$$\mathbb{E}_{\mathbb{X}}(\hat{T}_1) = \mathrm{Var}(\mathbb{E}(Y|X)) + M_1 h_1^2 + \frac{M_2}{nh_1} + o_P(h_1^2)$$

- Biais de $\hat{\sigma}^2(\cdot)$:

$$\mathbb{E}_{\mathbb{X}}(\hat{T}_2) = \mathbb{E}(\mathrm{Var}(Y|X)) + V_1 h_2^2 + o_P(h_1^2 + h_2^2)$$

- Variance de $\hat{\sigma}^2(\cdot)$:

$$\begin{aligned}\mathrm{Var}_{\mathbb{X}}(\hat{T}_2) &= \frac{1}{n'}\left\{\mathbb{E}(\mathrm{Var}(Y|X)^2) + V_2 h_2^2 + V_3 h_1^2 + \frac{V_4}{nh_2}\right. \\ &\quad \left. + o_P\left(h_1^2 + h_2^2 + \frac{1}{nh_2}\right)\right\}\end{aligned}$$