Estimation efficace des indices de Sobol du premier ordre

S. Da Veiga-F. G

IMT Toulouse

BoSanTouVal09
-
jeudi 04 Juin 2009


Colosseo
ATTENZIONE

Plan

Plan

Modélisation

- Phénomène physique, chimique, économique : modélisation
- Sorties, variables d'entrées, paramètres, etc.
- Confiance du modèle ?
  - Entrées incertaines
  - Paramètres incertains

Conséquence : incertitudes sur les sorties du modèle

Modèle :

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

avec Y sorties du modèle et $\mathbf{X} = (X_1, \ldots, X_d)$ vecteur des d facteurs (entrées + paramètres incertains).

Objectif

Quels sont les entrées et les paramètres qui ont le plus grand impact sur les sorties du modèle ?

Plan

Cadre probabiliste

Indices de sensibilité

$$S_i = \frac{\mathrm{Var}(\mathbb{E}(Y|X_i))}{\mathrm{Var}(Y)} \qquad \text{(1er ordre)}$$

- Interprétation :
  - $\mathbb{E}(Y|X_i)$: fonction de X_i uniquement qui *approche* le mieux Y
  - $\mathrm{Var}(\mathbb{E}(Y|X_i))$: fluctuation de la sortie si elle était fonction uniquement de X_i
  - puis normalisation par la fluctuation totale de Y, $\mathrm{Var}(Y)$

Cadre probabiliste

Indices de sensibilité

$$S_i = \frac{\mathrm{Var}(\mathbb{E}(Y|X_i))}{\mathrm{Var}(Y)} \qquad \text{(1er ordre)}$$

- Interprétation :
  - $\mathbb{E}(Y|X_i)$: fonction de X_i uniquement qui *approche* le mieux Y
  - $\mathrm{Var}(\mathbb{E}(Y|X_i))$: fluctuation de la sortie si elle était fonction uniquement de X_i
  - puis normalisation par la fluctuation totale de Y, $\mathrm{Var}(Y)$

Cadre probabiliste

- Techniques usuelles :
  - Sobol
  - FAST

Limitations

Nombre de runs
Hypothèse d'indépendance des entrées

- Exemples de cas problématiques
  - Réservoir pétrolier (temps de calcul)
  - Cinétique chimique (pas d'indépendance)

Cadre probabiliste

- Techniques usuelles :
  - Sobol
  - FAST

Limitations

Nombre de runs
Hypothèse d'indépendance des entrées

- Exemples de cas problématiques
  - Réservoir pétrolier (temps de calcul)
  - Cinétique chimique (pas d'indépendance)

Cadre probabiliste

- Techniques usuelles :
  - Sobol
  - FAST

Limitations

Nombre de runs
Hypothèse d'indépendance des entrées

- Exemples de cas problématiques
  - Réservoir pétrolier (temps de calcul)
  - Cinétique chimique (pas d'indépendance)

Cadre probabiliste

Sans hypothèse d'indépendance

- Méthode de Oakley et 0'Hagan :
  - $f(\mathbf{X})$ approchée par processus Gaussiens
  - Puis intégration / loi de $\mathbf{X}$

Evaluation numérique d'intégrales multiples (dimension $2d-1$)

- Méthode de Da-Veiga, Wahl et Gamboa :
  - $\mathbb{E}(Y|X_i)$ approchée par polynômes locaux
  - Plug-in pour $\mathrm{Var}(\mathbb{E}(Y|X_i))$

Vitesse de convergence non-paramétrique

Cadre probabiliste

Sans hypothèse d'indépendance

- Méthode de Oakley et 0'Hagan :
  - $f(\mathbf{X})$ approchée par processus Gaussiens
  - Puis intégration / loi de $\mathbf{X}$

Evaluation numérique d'intégrales multiples (dimension $2d-1$)

- Méthode de Da-Veiga, Wahl et Gamboa :
  - $\mathbb{E}(Y|X_i)$ approchée par polynômes locaux
  - Plug-in pour $\mathrm{Var}(\mathbb{E}(Y|X_i))$

Vitesse de convergence non-paramétrique

Cadre probabiliste

Sans hypothèse d'indépendance

- Méthode de Oakley et 0'Hagan :
  - $f(\mathbf{X})$ approchée par processus Gaussiens
  - Puis intégration / loi de $\mathbf{X}$

Evaluation numérique d'intégrales multiples (dimension $2d-1$)

- Méthode de Da-Veiga, Wahl et Gamboa :
  - $\mathbb{E}(Y|X_i)$ approchée par polynômes locaux
  - Plug-in pour $\mathrm{Var}(\mathbb{E}(Y|X_i))$

Vitesse de convergence non-paramétrique

Plan

Réécriture du problème

(X_i, Y_i), $i = 1, \ldots, n$ échantillon de (X, Y), densité f
On veut estimer

$$\mathrm{Var}(\mathbb{E}(Y|X)) = \mathbb{E}(\mathbb{E}(Y|X)^2) - (\mathbb{E}(Y))^2.$$

On définit l'opérateur T tel que

$$T(f) = \mathbb{E}(\mathbb{E}(Y|X)^2) = \iint \left(\frac{\int y f(x,y)dy}{\int f(x,y)dy} \right)^2 f(x,y)dxdy.$$

Travaux existants

B. Laurent (1996) a étudié l'estimation efficace de fonctionnelles de la forme

$$T(f) = \int \phi(f(\mathbf{x}), \mathbf{x}) d\mathbf{x}$$

où ϕ régulière et $X_1, \ldots, X_n$ de densité de probabilité f.

Idée de construction d'un estimateur

Développer $u \to \phi(u, x) \in C^3$ autour d'un estimateur $\hat{f}$ préliminaire

Travaux existants

B. Laurent (1996) a étudié l'estimation efficace de fonctionnelles de la forme

$$T(f) = \int \phi(f(\mathbf{x}), \mathbf{x}) d\mathbf{x}$$

où ϕ régulière et $X_1, \ldots, X_n$ de densité de probabilité f.

Idée de construction d'un estimateur

Développer $u \to \phi(u, x) \in C^3$ autour d'un estimateur $\hat{f}$ préliminaire

Travaux existants

B. Laurent (1996) a étudié l'estimation efficace de fonctionnelles de la forme

$$T(f) = \int \phi(f(\mathbf{x}), \mathbf{x}) d\mathbf{x}$$

où ϕ régulière et $X_1, \ldots, X_n$ de densité de probabilité f.

Idée de construction d'un estimateur

Développer $u \to \phi(u, x) \in C^3$ autour d'un estimateur $\hat{f}$ préliminaire

$$\begin{aligned}\phi(f(\mathbf{x}), \mathbf{x}) &= \phi(\hat{f}(\mathbf{x}), \mathbf{x}) + \frac{\partial \phi}{\partial u}(\hat{f}(\mathbf{x}), \mathbf{x}) \left(f(\mathbf{x}) - \hat{f}(\mathbf{x})\right) \\ &+ \frac{\partial^2 \phi}{\partial u^2}(\hat{f}(\mathbf{x}), \mathbf{x}) \left(f(\mathbf{x}) - \hat{f}(\mathbf{x})\right)^2 + \gamma.\end{aligned}$$

Travaux existants

B. Laurent (1996) a étudié l'estimation efficace de fonctionnelles de la forme

$$T(f) = \int \phi(f(\mathbf{x}), \mathbf{x}) d\mathbf{x}$$

où ϕ régulière et $X_1, \ldots, X_n$ de densité de probabilité f.

Idée de construction d'un estimateur

Développer $u \to \phi(u, x) \in C^3$ autour d'un estimateur $\hat{f}$ préliminaire

$$\begin{aligned}\phi(f(\mathbf{x}), \mathbf{x}) &= \phi(\hat{f}(\mathbf{x}), \mathbf{x}) + \phi_1'(\hat{f}(\mathbf{x}), \mathbf{x}) \left(f(\mathbf{x}) - \hat{f}(\mathbf{x})\right) \\ &\quad + \phi_1''(\hat{f}(\mathbf{x}), \mathbf{x}) \left(f(\mathbf{x}) - \hat{f}(\mathbf{x})\right)^2 + \gamma.\end{aligned}$$

Travail de B. Laurent

$$T(f) = \int \phi(f(\mathbf{x}), \mathbf{x}) d\mathbf{x}$$

- Γ_n négligeable
- $\int H(\hat{f}, \cdot)\, f$ fonctionnelle linéaire de f : pas de problème
- $\int K(\hat{f}, \cdot)\, f^2$ fonctionnelle quadratique de f : estimation optimale (efficacité) ?

Travail de B. Laurent

$$\begin{aligned} T(f) &= \int \phi(f(\mathbf{x}),\mathbf{x})d\mathbf{x} \\ &= \int \phi(\hat{f}(\mathbf{x}),\mathbf{x})d\mathbf{x} + \int \phi_1'(\hat{f}(\mathbf{x}),\mathbf{x}) \left(f(\mathbf{x}) - \hat{f}(\mathbf{x})\right) d\mathbf{x} \\ &+ \int \phi_1''(\hat{f}(\mathbf{x}),\mathbf{x}) \left(f(\mathbf{x}) - \hat{f}(\mathbf{x})\right)^2 d\mathbf{x} + \Gamma_n \end{aligned}$$

- Γ_n négligeable
- $\displaystyle\int H(\hat{f},\cdot)\, f$ fonctionnelle linéaire de f : pas de problème
- $\displaystyle\int K(\hat{f},\cdot)\, f^2$ fonctionnelle quadratique de f : estimation optimale (efficacité) ?

Travail de B. Laurent

$$\begin{aligned} T(f) &= \int \phi(f(\mathbf{x}), \mathbf{x}) d\mathbf{x} \\ &= \int \phi(\hat{f}(\mathbf{x}), \mathbf{x}) d\mathbf{x} + \int \phi_1'(\hat{f}(\mathbf{x}), \mathbf{x}) \left(f(\mathbf{x}) - \hat{f}(\mathbf{x}) \right) d\mathbf{x} \\ &\quad + \int \phi_1''(\hat{f}(\mathbf{x}), \mathbf{x}) \left(f(\mathbf{x}) - \hat{f}(\mathbf{x}) \right)^2 d\mathbf{x} + \Gamma_n \\ &= \int G(\hat{f}, \cdot) + \int H(\hat{f}, \cdot) f + \int K(\hat{f}, \cdot) f^2 + \Gamma_n \end{aligned}$$

- Γ_n négligeable
- $\int H(\hat{f}, \cdot) f$ fonctionnelle linéaire de f : pas de problème
- $\int K(\hat{f}, \cdot) f^2$ fonctionnelle quadratique de f : estimation optimale (efficacité) ?

Travail de B. Laurent

$$
\begin{aligned}
T(f) &= \int \phi(f(\mathbf{x}),\mathbf{x})d\mathbf{x} \\
&= \int \phi(\hat{f}(\mathbf{x}),\mathbf{x})d\mathbf{x} + \int \phi_1'(\hat{f}(\mathbf{x}),\mathbf{x})\left(f(\mathbf{x}) - \hat{f}(\mathbf{x})\right)d\mathbf{x} \\
&\quad + \int \phi_1''(\hat{f}(\mathbf{x}),\mathbf{x})\left(f(\mathbf{x}) - \hat{f}(\mathbf{x})\right)^2 d\mathbf{x} + \Gamma_n \\
&= \int G(\hat{f},\cdot) + \int H(\hat{f},\cdot)\,f + \int K(\hat{f},\cdot)\,f^2 + \Gamma_n
\end{aligned}
$$

- Γ_n négligeable
- $\int H(\hat{f},\cdot)\,f$ fonctionnelle linéaire de f : pas de problème
- $\int K(\hat{f},\cdot)\,f^2$ fonctionnelle quadratique de f : estimation optimale (efficacité) ?

Travail de B. Laurent

$$
\begin{aligned}
T(f) &= \int \phi(f(\mathbf{x}), \mathbf{x}) d\mathbf{x} \\
&= \int \phi(\hat{f}(\mathbf{x}), \mathbf{x}) d\mathbf{x} + \int \phi_1'(\hat{f}(\mathbf{x}), \mathbf{x}) \left(f(\mathbf{x}) - \hat{f}(\mathbf{x}) \right) d\mathbf{x} \\
&+ \int \phi_1''(\hat{f}(\mathbf{x}), \mathbf{x}) \left(f(\mathbf{x}) - \hat{f}(\mathbf{x}) \right)^2 d\mathbf{x} + \Gamma_n \\
&= \int G(\hat{f}, \cdot) + \int H(\hat{f}, \cdot) f + \int K(\hat{f}, \cdot) f^2 + \Gamma_n
\end{aligned}
$$

- Γ_n négligeable
- $\int H(\hat{f}, \cdot) f$ fonctionnelle linéaire de f : pas de problème
- $\int K(\hat{f}, \cdot) f^2$ fonctionnelle quadratique de f : estimation optimale (efficacité) ?

Travail de B. Laurent

$$
\begin{aligned}
T(f) &= \int \phi(f(\mathbf{x}), \mathbf{x}) d\mathbf{x} \\
&= \int \phi(\hat{f}(\mathbf{x}), \mathbf{x}) d\mathbf{x} + \int \phi_1'(\hat{f}(\mathbf{x}), \mathbf{x}) \left(f(\mathbf{x}) - \hat{f}(\mathbf{x}) \right) d\mathbf{x} \\
&+ \int \phi_1''(\hat{f}(\mathbf{x}), \mathbf{x}) \left(f(\mathbf{x}) - \hat{f}(\mathbf{x}) \right)^2 d\mathbf{x} + \Gamma_n \\
&= \int G(\hat{f}, \cdot) + \int H(\hat{f}, \cdot) f + \int K(\hat{f}, \cdot) f^2 + \Gamma_n
\end{aligned}
$$

- Γ_n négligeable
- $\int H(\hat{f}, \cdot) f$ fonctionnelle linéaire de f : pas de problème
- $\int K(\hat{f}, \cdot) f^2$ fonctionnelle quadratique de f : estimation optimale (efficacité) ?

Estimation efficace de fonctionnelles quadratique

On veut estimer
$$\int K(\hat{f},\cdot)\, f^2 \equiv \int \psi\, f^2$$
où $f \in \mathbb{L}^2(dx)$ densité de X, échantillon $(X_1, \ldots, X_n)$.

Estimation efficace de fonctionnelles quadratique

On veut estimer

$$\int K(\hat{f}, \cdot)\, f^2 \equiv \int \psi\, f^2$$

où $f \in \mathbb{L}^2(dx)$ densité de X, échantillon $(X_1, \ldots, X_n)$.

Soit $(p_i)_{i \in D}$ base orthonormale de $\mathbb{L}^2(dx)$, D dénombrable, $a_i = \int f p_i$.

Estimation par projection

Estimation efficace de fonctionnelles quadratique

On veut estimer

$$\int K(\hat{f},\cdot)\, f^2 \equiv \int \psi f^2$$

où $f \in \mathbb{L}^2(dx)$ densité de X, échantillon $(X_1,\ldots,X_n)$.

Soit $(p_i)_{i\in D}$ base orthonormale de $\mathbb{L}^2(dx)$, D dénombrable, $a_i = \int fp_i$.

Estimation par projection

Cas particulier : $\psi = 1$

- $\displaystyle \int f^2 = \sum_{i\in D} a_i^2 \Rightarrow \tilde{\theta} = \sum_{i\in M} \hat{a}_i^2,\ \hat{a}_i = \frac{1}{n}\sum_{j=1}^{n} p_i(X_j),\ M \subset D$
- Réduction du biais : $\displaystyle \tilde{\theta} = \frac{1}{n(n-1)} \sum_{i\in M} \sum_{j\neq k=1}^{n} p_i(X_j)p_i(X_k)$

Estimation efficace de fonctionnelles quadratique

$$\psi = 1 \qquad \tilde{\theta} = \frac{1}{n(n-1)} \sum_{i \in M} \sum_{j \neq k = 1}^{n} p_i(X_j) p_i(X_k)$$

Estimation efficace de fonctionnelles quadratique

$$\psi = 1 \qquad \tilde{\theta} = \frac{1}{n(n-1)} \sum_{i \in M} \sum_{j \neq k = 1}^{n} p_i(X_j) p_i(X_k)$$

$$\text{Biais}(\tilde{\theta}) = -\int (S_M f - f)^2 \quad \text{avec} \quad S_M f = \sum_{i \in M} a_i p_i$$

Seulement un bias de projection

Estimation efficace de fonctionnelles quadratique

$$\psi = 1 \qquad \tilde{\theta} = \frac{1}{n(n-1)} \sum_{i \in M} \sum_{j \neq k = 1}^{n} p_i(X_j) p_i(X_k)$$

$$\mathrm{Biais}(\tilde{\theta}) = -\int (S_M f - f)^2 \quad \text{avec} \quad S_M f = \sum_{i \in M} a_i p_i$$

Seulement un bias de projection

ψ quelconque : $-\int (S_M f - f)^2 \psi = 2 \int (S_M f) f \psi - \int (S_M f)^2 \psi - \int \psi f^2$

Estimation efficace de fonctionnelles quadratique

$$\psi = 1 \qquad \tilde{\theta} = \frac{1}{n(n-1)} \sum_{i \in M} \sum_{j \neq k = 1}^{n} p_i(X_j) p_i(X_k)$$

$$\mathrm{Biais}(\tilde{\theta}) = -\int (S_M f - f)^2 \quad \text{avec} \quad S_M f = \sum_{i \in M} a_i p_i$$

Seulement un bias de projection

ψ quelconque : $-\int (S_M f - f)^2 \psi = 2\int (S_M f) f \psi - \int (S_M f)^2 \psi - \int \psi f^2$

Estimateur proposé :

$$\begin{aligned} \hat{\theta} &= \frac{2}{n(n-1)} \sum_{i \in M} \sum_{j \neq k = 1}^{n} p_i(X_j)(p_i \psi)(X_k) \\ &\quad - \frac{1}{n(n-1)} \sum_{i, i' \in M} \sum_{j \neq k = 1}^{n} p_i(X_j) p_{i'}(X_k) \int p_i p_{i'} \psi(x) dx \end{aligned}$$

Estimation efficace de fonctionnelles quadratique

Hypothèses

$X_1, \ldots, X_n$ i.i.d. $f \in \mathbb{L}^2(dx)$. $(p_i)_{i \in D}$ base orthonormée de $\mathbb{L}^2(dx)$

- ψ bornée, f uniformément bornée et $f \in \mathcal{E}$,
  $$\mathcal{E} = \left\{ \sum_{i \in D} a_i p_i ; \sum_{i \in D} \left| \frac{a_i^2}{c_i^2} \right| \leq 1 \right\}$$
- $M_n \subset D$ tq $\left(\sup_{i \notin M_n} |c_i|^2 \right)^2 \approx \frac{|M_n|}{n^2}$
- $\forall t \in \mathbb{L}^2(d\mu)$, $\int (S_{M_n} t - t)^2 d\mu \to 0$

Estimation efficace de fonctionnelles quadratique

Théorème (Estimation efficace de $\theta = \int \psi f^2$)

(i) Si $|M|/n \to 0$

$$\sqrt{n}\left(\hat{\theta} - \theta\right) \to \mathcal{N}\left(0, \Lambda(f,\psi)\right),$$

$$\left|\mathbb{E}\left(\hat{\theta} - \theta\right)^2 - \Lambda(f,\psi)\right| \leq \gamma_1 \left[\frac{|M|}{n} + \|S_M f - f\|_2 + \|S_M(f\psi) - f\psi\|_2\right]$$

$$\text{où} \quad \Lambda(f,\psi) = 4\left[\int f^3\psi^2 - \left(\int f^2\psi\right)^2\right],$$

(ii) Sinon

$$\mathbb{E}\left(\hat{\theta} - \theta\right)^2 \leq \gamma_2 \frac{|M|}{n}.$$

Estimation efficace de fonctionnelles quadratique

Théorème (Estimation efficace de $\theta = \int \psi f^2$)

(i) Si $|M|/n \to 0$

$$\sqrt{n}\left(\hat{\theta} - \theta\right) \to \mathcal{N}\left(0, \Lambda(f,\psi)\right),$$

$$\left|\mathbb{E}\left(\hat{\theta} - \theta\right)^2 - \Lambda(f,\psi)\right| \leq \gamma_1 \left[\frac{|M|}{n} + \|S_M f - f\|_2 + \|S_M(f\psi) - f\psi\|_2\right]$$

$$\text{où} \quad \Lambda(f,\psi) = 4\left[\int f^3\psi^2 - \left(\int f^2\psi\right)^2\right],$$

(ii) Sinon

$$\mathbb{E}\left(\hat{\theta} - \theta\right)^2 \leq \gamma_2 \frac{|M|}{n}.$$

$$\lim_{n\to\infty} n\mathbb{E}\left(\hat{\theta}_n - \theta\right)^2 = \Lambda(f,\psi)$$

Estimation efficace de fonctionnelles quadratique

Théorème (Estimation efficace de $\theta = \int \psi f^2$)

(i) Si $|M|/n \to 0$

$$\sqrt{n}\left(\hat\theta - \theta\right) \to \mathcal{N}\left(0, \Lambda(f,\psi)\right),$$

$$\left|\mathbb{E}\left(\hat\theta - \theta\right)^2 - \Lambda(f,\psi)\right| \leq \gamma_1 \left[\frac{|M|}{n} + \|S_M f - f\|_2 + \|S_M(f\psi) - f\psi\|_2\right]$$

$$\text{où} \quad \Lambda(f,\psi) = 4\left[\int f^3\psi^2 - \left(\int f^2\psi\right)^2\right],$$

(ii) Sinon

$$\mathbb{E}\left(\hat\theta - \theta\right)^2 \leq \gamma_2 \frac{|M|}{n}.$$

$\Lambda(f,\psi)$ optimale (borne de Cramér-Rao semi-paramétrique)

Théorème (Estimation efficace de $\theta = \int \psi f^2$)

(i) Si $|M|/n \to 0$

$$\sqrt{n}\left(\hat{\theta} - \theta\right) \to \mathcal{N}\left(0, \Lambda(f,\psi)\right),$$

$$\left|\mathbb{E}\left(\hat{\theta} - \theta\right)^2 - \Lambda(f,\psi)\right| \leq \gamma_1 \left[\frac{|M|}{n} + \|S_M f - f\|_2 + \|S_M(f\psi) - f\psi\|_2\right]$$

$$\text{où} \quad \Lambda(f,\psi) = 4\left[\int f^3\psi^2 - \left(\int f^2\psi\right)^2\right],$$

(ii) Sinon

$$\mathbb{E}\left(\hat{\theta} - \theta\right)^2 \leq \gamma_2 \frac{|M|}{n}.$$

Bornes explicites car ψ aléatoire et dépendant de n dans Taylor

Estimation efficace de fonctionnelles quadratique

Esquisse de la démonstration

$$\mathbb{E}\left(\hat{\theta} - \int \psi f^2\right)^2 = \text{Biais}^2(\hat{\theta}) + \text{Var}(\hat{\theta})$$

- $\text{Biais}(\hat{\theta}) = -\int (S_M f - f)^2 \psi$:

$$\left|\text{Biais}(\hat{\theta})\right| \leq \|\psi\|_\infty \int (S_M f - f)^2 = \|\psi\|_\infty \sum_{i \neq M} |a_i|^2 \leq \|\psi\|_\infty \sup_{i \neq M} |c_i|^2$$

- Partie technique : majoration de $\text{Var}(\hat{\theta})$

Retour au dvt de Taylor

$$T(f) = \int G(\hat{f},\cdot) + \int \left(\phi_1'(\hat{f},\cdot) - \hat{f}\phi_1''(\hat{f},\cdot)\right) f + \int \frac{1}{2}\phi_1''(\hat{f},\cdot) f^2 + \Gamma_n$$

estimé par

$$\begin{aligned}
\hat{T}_n &= \int G(\hat{f},\cdot) + \frac{1}{n_2}\sum_{j=1}^{n_2}\left(\phi_1'(\hat{f},\cdot) - \hat{f}\phi_1''(\hat{f},\cdot)\right)(X_j) \\
&\quad + \frac{2}{n_2(n_2-1)}\sum_{i\in M}\sum_{j\neq k=1}^{n_2} p_i(X_j)(p_i\frac{1}{2}\phi_1''(\hat{f},\cdot))(X_k) \\
&\quad - \frac{1}{n_2(n_2-1)}\sum_{i,i'\in M}\sum_{j\neq k=1}^{n_2} p_i(X_j)p_{i'}(X_k)\int p_i p_{i'}\frac{1}{2}\phi_1''(\hat{f},\cdot)
\end{aligned}$$

Retour au dvt de Taylor

On montre que $\hat{T}_n - T(f)$ a même comportement asymptotique que

$$\frac{1}{n_2}\sum_{j=1}^{n_2} \phi_1'(f,\cdot)(X_j) - \int \phi_1'(f,\cdot) f$$

Convergence de $\hat{T}_n$

Hypothèses

$X_1, \ldots, X_n$ i.i.d. $f \in \mathbb{L}^2(dx)$. $(p_i)_{i \in D}$ base orthonormée de $\mathbb{L}^2(dx)$

- $\phi \in C^3$, f uniformément bornée et $f \in \mathcal{E}$
- $a \leq f \leq b$ et $a - \epsilon \leq \hat{f} \leq b + \epsilon$
- $\mathbb{E}\|\hat{f} - f\|_q^l \leq C n_1^{-l\lambda}$ $\forall l,\ q \geq 2,\ \lambda > 1/6$
- $M_n \subset D$ tq $\left(\sup_{i \notin M_n} |c_i|^2\right)^2 \approx \dfrac{|M_n|}{n^2}$
- $\forall t \in \mathbb{L}^2(d\mu),\ \int (S_{M_n} t - t)^2 d\mu \to 0$

Conclusion travail de B. Laurent

Théorème (Estimation efficace de $T(f)$)

(i) Si $|M|/n \to 0$
$$\sqrt{n}\left(\hat{T}_n - T(f)\right) \to \mathcal{N}(0, C(f,\phi))$$

(ii)
$$\lim_{n\to\infty} n\mathbb{E}(\hat{T}_n - T(f))^2 = C(f,\phi)$$

où
$$C(f,\phi) = \int (\phi_1'(f,\cdot))^2 f - \left(\int \phi_1'(f,\cdot) f\right)^2$$

$C(f,\phi)$ constante optimale

Retour à l'estimation des indices de sensibilité

On veut estimer

$$T(f) = \mathbb{E}(\mathbb{E}(Y|X)^2) = \iint \left(\frac{\int y\, f(x,y) dy}{\int f(x,y) dy} \right)^2 f(x,y) dx dy.$$

On suit la méthode précédente : on développe $T(f)$ autour d'un estimateur préliminaire $\hat{f}$.

Développement de $T(f)$

$$
\begin{aligned}
T(f) &= \iint \left[2y\hat{m}(x) - \hat{m}(x)^2\right] f(x,y)dxdy \\
&+ \iiint \frac{1}{\left(\int \hat{f}(x,y)dy\right)} \left[yz + \hat{m}(x)^2 - (y+z)\hat{m}(x)\right] f(x,y)f(x,z)dxdydz \\
&+ \Gamma_n \\
&= \iint H(\hat{f},x,y)f(x,y)dxdy + \iiint K(\hat{f},x,y,z)f(x,y)f(x,z)dxdydz \\
&+ \Gamma_n
\end{aligned}
$$

où

$$
\begin{aligned}
H(\hat{f},x,y) &= 2y\hat{m}(x) - \hat{m}(x)^2 \\
K(\hat{f},x,y,z) &= \frac{1}{\left(\int \hat{f}(x,y)dy\right)} \left[yz + \hat{m}(x)^2 - (y+z)\hat{m}(x)\right].
\end{aligned}
$$

Estimation de fonctionnelles quadratiques

On veut estimer

$$\iiint K(\hat{f}, x, y, z) f(x, y) f(x, z) = \iiint \psi(x, y, z) f(x, y) f(x, z)$$

- Idée : même procédure que B. Laurent
- On cherche un estimateur de biais

$$\begin{aligned}
& - \iiint [S_M f(x, y) - f(x, y)][S_M f(x, z) - f(x, z)] \psi(x, y, z) dxdydz \\
= \; & 2 \iiint S_M f(x, y) f(x, z) \psi(x, y, z) dxdydz \\
& - \iiint S_M f(x, y) S_M f(x, z) \psi(x, y, z) dxdydz \\
& - \iiint f(x, y) f(x, z) \psi(x, y, z) dxdydz
\end{aligned}$$

Estimation de fonctionnelles quadratiques

Estimateur proposé :

$$
\begin{aligned}
\hat{\theta}_n &= \frac{2}{n(n-1)} \sum_{i \in M} \sum_{j \neq k = 1}^{n} p_i(X_j, Y_j) \int p_i(X_k, u) \psi(X_k, u, Y_k) du \\
&\quad - \frac{1}{n(n-1)} \sum_{i, i' \in M} \sum_{j \neq k = 1}^{n} p_i(X_j, Y_j) p_{i'}(X_k, Y_k) \\
&\quad \iiint p_i(x, y) p_{i'}(x, z) \psi(x, y, z) dx dy dz.
\end{aligned}
$$

Hypothèses

$X_1, \ldots, X_n$ i.i.d. $f \in \mathbb{L}^2(dx)$. $(p_i)_{i \in D}$ base orthonormée de $\mathbb{L}^2(dx)$

- $\psi(x,y,z)$ bornée, f uniformément bornée et $f \in \mathcal{E}$,
  $$\mathcal{E} = \left\{ \sum_{i \in D} a_i p_i ; \sum_{i \in D} \left| \frac{a_i^2}{c_i^2} \right| \leq 1 \right\}$$
- $M_n \subset D$ tq $\left(\sup_{i \notin M_n} |c_i|^2 \right)^2 \approx \frac{|M_n|}{n^2}$
- $\forall t \in \mathbb{L}^2(d\mu)$, $\int (S_{M_n} t - t)^2 d\mu \to 0$

Théorème (Estimation efficace de $\theta = \int \psi(x,y,z)f(x,y)f(x,z)$)

(i) Si $|M|/n \to 0$

$$\sqrt{n}\left(\hat{\theta}-\theta\right) \to \mathcal{N}\left(0, \Lambda(f,\psi)\right),$$

$$\left|\mathbb{E}\left(\hat{\theta}-\theta\right)^2 - \Lambda(f,\psi)\right| \leq \gamma_1 \left[\frac{|M|}{n} + \|S_M f - f\|_2 + \|S_M(g) - g\|_2\right]$$

$$\text{où} \quad \Lambda(f,\psi) = 4\left[\iint g(x,y)^2 f(x,y)dxdy - \left(\iint g(x,y)f(x,y)dxdy\right)^2\right]$$

où

$$g(x,y) = \int f(x,u)\psi(x,y,u)du$$

Retour au dvt de Taylor

$$\begin{aligned} T(f) &= \iint H(\hat{f}, x, y) f(x, y)\, dxdy + \iiint K(\hat{f}, x, y, z) f(x, y) f(x, z)\, dxdydz \\ &+ \Gamma_n \end{aligned}$$

Estimateur proposé :

$$\begin{aligned} \hat{\theta}_n &= \frac{1}{n_2} \sum_{j=1}^{n_2} H(\hat{f}, X_j, Y_j) \\ &+ \frac{2}{n_2(n_2-1)} \sum_{i \in M} \sum_{j \neq k = 1}^{n_2} p_i(X_j, Y_j) \int p_i(X_k, u) K(\hat{f}, X_k, u, Y_k)\, du \\ &- \frac{1}{n_2(n_2-1)} \sum_{i, i' \in M} \sum_{j \neq k = 1}^{n_2} p_i(X_j, Y_j) p_{i'}(X_k, Y_k) \\ &\iiint p_i(x, y) p_{i'}(x, z) K(\hat{f}, x, y, z)\, dxdydz. \end{aligned}$$

Propriétés de $\hat{T}_n$

Théorème (Estimation efficace de $T(f)$)

(i) Si $|M|/n \to 0$

$$\sqrt{n}\left(\hat{T}_n - T(f)\right) \to \mathcal{N}(0, C(f))$$

(ii)

$$\lim_{n\to\infty} n\mathbb{E}(\hat{T}_n - T(f))^2 = C(f)$$

où

$$C(f) = \iint H(f,x,y)^2 f(x,y)d\mu(x,y) - \left(\iint H(f,x,y)f(x,y)d\mu(x,y)\right)^2$$

$C(f) = 4\mathbb{E}\left(\mathrm{Var}(Y|X)\mathbb{E}(Y|X)^2\right) + \mathrm{Var}\left(\mathbb{E}(Y|X)^2\right)$ constante optimale

Extensions

- On peut généraliser très facilement ce résultat aux estimateurs des fonctionnelles

$$\mathbb{E}\Big(\psi\big(\mathbb{E}(\phi(Y)|X)\big)\Big)$$

si $f \in \mathbb{L}^2(dxdy)$, $\phi \in C^0$ et $\psi \in C^3$.

- Constante optimale :

$$C(f) = \mathbb{E}\left(\mathrm{Var}(\phi(Y)|X)\left[\dot{\psi}\left(\mathbb{E}(Y|X)\right)\right]^2\right) + \mathrm{Var}\left(\psi\left(\mathbb{E}(\phi(Y)|X)\right)\right)$$

Meilleure procédure d'estimation

- Pour estimer $\mathbb{E}(\mathbb{E}(Y|X^i)^2)$, on utilise un échantillon du couple (X^i, Y)
- On ne tient pas compte des réalisations des autres facteurs X^j, $j \neq i$

Perd-on de l'information ?

Asymptotiquement, les autres composantes n'apportent pas d'information pour estimer S_i

Meilleure procédure d'estimation

- Pour estimer $\mathbb{E}(\mathbb{E}(Y|X^i)^2)$, on utilise un échantillon du couple (X^i, Y)
- On ne tient pas compte des réalisations des autres facteurs X^j, $j \neq i$

Perd-on de l'information ?

Asymptotiquement, les autres composantes n'apportent pas d'information pour estimer S_i

Meilleure procédure d'estimation

- Pour estimer $\mathbb{E}(\mathbb{E}(Y|X^i)^2)$, on utilise un échantillon du couple (X^i, Y)
- On ne tient pas compte des réalisations des autres facteurs X^j, $j \neq i$

Perd-on de l'information ?

Réponse :

- La constante optimale en considérant toutes les composantes $(X^1, \ldots, X^d)$ est encore $C(f)$
- Or notre estimateur $\widehat{T}_n$ atteint cette vitesse en travaillant uniquement avec (X^i, Y)

Asymptotiquement, les autres composantes n'apportent pas d'information pour estimer S_i

Meilleure procédure d'estimation

- Pour estimer $\mathbb{E}(\mathbb{E}(Y|X^i)^2)$, on utilise un échantillon du couple (X^i, Y)
- On ne tient pas compte des réalisations des autres facteurs X^j, $j \neq i$

Perd-on de l'information ?

Réponse :

- La constante optimale en considérant toutes les composantes $(X^1, \ldots, X^d)$ est encore $C(f)$
- Or notre estimateur $\widehat{T}_n$ atteint cette vitesse en travaillant uniquement avec (X^i, Y)

Asymptotiquement, les autres composantes n'apportent pas d'information pour estimer S_i

Meilleure procédure d'estimation

- Pour estimer $\mathbb{E}(\mathbb{E}(Y|X^i)^2)$, on utilise un échantillon du couple (X^i, Y)
- On ne tient pas compte des réalisations des autres facteurs X^j, $j \neq i$

Perd-on de l'information ?

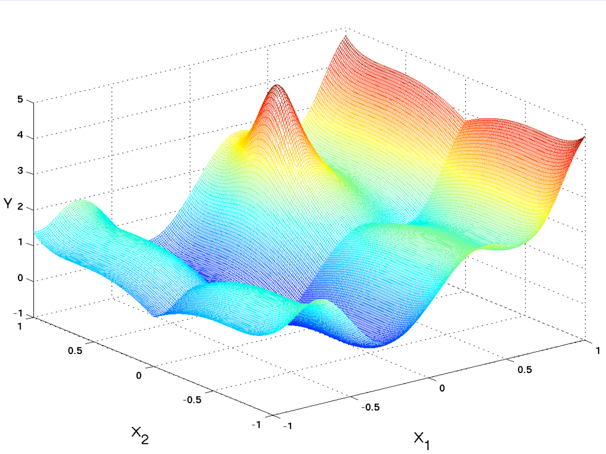
Réponse :

- La constante optimale en considérant toutes les composantes $(X^1, \ldots, X^d)$ est encore $C(f)$
- Or notre estimateur $\widehat{T}_n$ atteint cette vitesse en travaillant uniquement avec (X^i, Y)

Asymptotiquement, les autres composantes n'apportent pas d'information pour estimer S_i

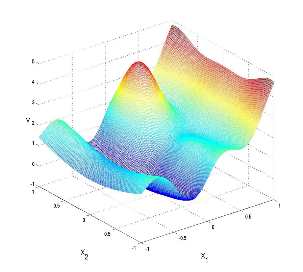
Exemple analytique

$$
\begin{aligned}
Y \quad = \quad & 0.2\exp(X_1 - 3) + 2.2|X_2| + 1.3X_2^6 - 2X_2^2 - 0.5X_2^4 - 0.5X_1^4 \\
& + 2.5X_1^2 + 0.7X_1^3 + \frac{3}{(8X_1 - 2)^2 + (5X_2 - 3)^2 + 1} + \sin(5X_1)\cos(3X_1^2)
\end{aligned}
$$

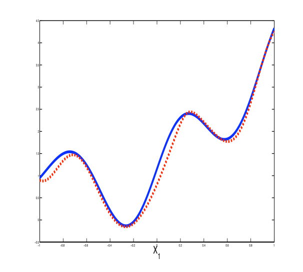


Krigeage (courbe théorique, approximation)

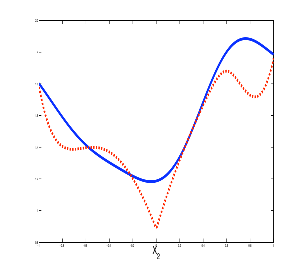
$f(X_1, X_2)$



$\mathbb{E}(Y|X_1)$

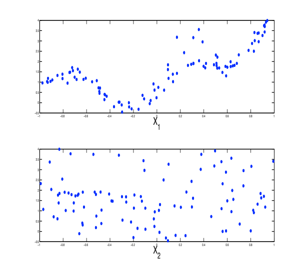


$\mathbb{E}(Y|X_2)$

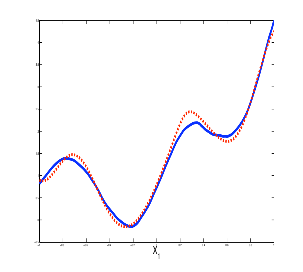


Polynômes locaux (courbe théorique, approximation)

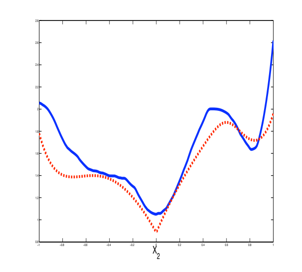
Échantillons marginaux



$\mathbb{E}(Y|X_1)$



$\mathbb{E}(Y|X_2)$



Exemple analytique

		Krigeage	Poly. loc.	Est. eff.
		100 pts	100 pts	100 pts
$\mathrm{Var}(\mathbb{E}(Y\|X_1))$	1.0932	1.0539	1.0643	1.1701
$\mathrm{Var}(\mathbb{E}(Y\|X_2))$	0.0729	0.1121	0.0527	0.0939

X_1 : résultats comparables

X_2 : meilleure précision pour les approximations marginales

Perspectives

- Tests numériques (efficacité asymptotique, quid avec n petit ?)
- Autres fonctionnelles faisant intervenir ces formes quadratiques ?
- Méthodes adaptatives ?

Muchas Gracias