Estimation d'un méta-modèle et des indices de sensibilité

S. Huet

INRA. MaIAGE. Jouy-en-Josas

30 août 2016

En collaboration avec M.L. Taupin, Université Evry Val-d'Essonne

1. Méta-modèle pour les variations du pH dans le lait
2. Estimation de la décomposition de Hoeffding-Sobol, régression gaussienne
3. Majoration du risque de l'estimateur du méta-modèle
4. Décomposition de type ANOVA sur des RKHS
5. Calcul de l'estimateur
6. Estimation des indices de Sobol
7. Retour aux variations du pH dans le lait
8. Questions ouvertes

Méta-modèle pour les variations du pH dans le lait

- Industrie laitière : comment l'acidification du lait modifie ses capacités à la fabrication du fromage
- Quel mécanisme sous-jacent à la digestion du lait ?

Un modèle, basé sur des réactions chimiques, décrit la répartition des ions dans différentes phases du lait : $\approx$ 100 composants y participent

- déterminer les constituants du lait qui agissent sur les variations du pH
- approcher le modèle par un modèle simplifié, dont le calcul est rapide
  ex : en vue de son intégration dans la modélisation de la dynamique de digestion des protéines laitières chez le porc (B. Laroche, MaIAGE)
- 4 variables explicatives : $\mathbf{X}$
  - Volume total : lait + HCl ajouté
  - Concentration en ions H^+ ajoutés
  - Concentration de ions calcium ajoutés
  - Concentration de caséine ajoutée

$$
\begin{array}{ccc}
\mathbf{X} \longrightarrow & \boxed{\text{Expérimentation}} & \longrightarrow Y(\mathbf{X}) = \phi(\mathbf{X}) + \varepsilon \\
& \updownarrow & \\
\mathbf{X}\,(+\ \text{Env.}) & \boxed{\text{Phénomène}} & \longrightarrow \phi(\mathbf{X},\ \text{Env.}) \approx \phi(\mathbf{X}) +\ \text{EffEnv} \\
& \updownarrow & \\
\mathbf{X} \longrightarrow & \boxed{\text{Modèle numérique}} & \longrightarrow m(\mathbf{X}) = \phi(\mathbf{X}) + \varepsilon
\end{array}
$$

Modèle non-paramétrique en régression

Modèle statistique gaussien :

$$Y_i = \phi(\mathbf{X}_i) + \varepsilon_i, i = 1, \ldots, n$$

- $\varepsilon_1, \ldots \varepsilon_n$ i.i.d $\mathcal{N}(0, \sigma^2)$
- $\mathbf{X}_i = (X_{i1}, \ldots, X_{id})$ i.i.d. de loi $\mathbb{P} = \mathbb{P}_1 \times \ldots \times \mathbb{P}_d$ sur $\mathcal{X}^d \subset \mathbb{R}^d$
  Les v.a. $X_1, \ldots, X_d$ indépendantes, X_a de loi $\mathbb{P}_a$, connue
- $\phi : \mathcal{X} \subset \mathbb{R}^d \to \mathbb{R}$ inconnue

Objectif : Estimer ϕ, tenant compte

- d grand
- fortes non-linéarités
- effets d'interactions entre les variables $X_1, \ldots, X_d$.

Modèle non-paramétrique en régression

Modèle additif :

$$\begin{aligned}\phi(\mathbf{X}) &\approx f(\mathbf{X}) = \mu + \sum_{a=1}^{d} f_a(X_a) \text{ avec } f_a(x) = \sum_j \theta_{aj}\psi_{aj}(x) \\ &\approx f(\mathbf{X}) = \sum_{a=1}^{d} \theta_a^T \Psi_a(X_a)\end{aligned}$$

par exemple : linéaire par morceaux : $f_a(x) = \sum_{j\in J_a} \theta_{aj} \min(x, z_j)$, $z_1 < z_2 < \ldots$.

- ne permet pas d'estimer les effets d'interaction, par exemple $f_{a_1 a_2}(X_{a_1}, X_{a_2})$
- objectif : estimer la décomposition de Hoeffding-Sobol de ϕ

Notations :

- $\mathcal{P}$ l'ensemble des parties de $\{1, \ldots, d\}$:

$$\mathcal{P} = \{\{1\}, \{2\}, \ldots, \{1,2\}, \{1,3\}, \ldots, \{1,2,3\} \ldots\}$$

- $\mathcal{X} = \mathcal{X}_1 \times \ldots \times \mathcal{X}_d$, $\mathcal{X}_a \subset \mathbb{R}$. $f : \mathcal{X} \to \mathbb{R}$
- $v \in \mathcal{P}$, $\mathbf{X}_v = (X_a, a \in v)$, $f_v : \mathcal{X}_v \to \mathbb{R}$

$$\text{si } v = \{1,2\}, \mathbf{X}_v = (X_1, X_2), f_v : \mathcal{X}_1 \times \mathcal{X}_2 \to \mathbb{R}$$

Décomposition de Hoeffding-Sobol

Sobol, I. M. (1993), Van Der Vaart, A. W. (2000), p. 157

- Si $\mathbf{X} \sim \mathbb{P} = \mathbb{P}_1 \times \ldots \times \mathbb{P}_d$

$$\begin{aligned}\phi(\mathbf{X}) &= \phi_0 + \sum_{a=1}^{d} \phi_a(X_a) + \sum_{a_1,a_2=1}^{d} \phi_{a_1 a_2}(X_{a_1}, X_{a_2}) + \ldots \\ &= \phi_0 + \sum_{v \in \mathcal{P}(d)} \phi_v(\mathbf{X}_v)\end{aligned}$$

  avec $E_{\mathbb{P}} \phi_v(\mathbf{X}_v) = 0$ et $E_{\mathbb{P}} \phi_v(\mathbf{X}_v) \phi_{v'}(\mathbf{X}_{v'}) = 0$ si $v \neq v'$.
- cette décomposition existe et est unique
- $\rightsquigarrow$ interprétation des effets des variables $\mathbf{X}_v$ sur la réponse ϕ
- $\rightsquigarrow$ décomposition de la variance

$$\mathbb{V}\left(\phi(\mathbf{X})\right) = \sum_{v \in \mathcal{P}(d)} \mathbb{V}\left(\phi_v(\mathbf{X}_v)\right)$$

- $\rightsquigarrow$ indices de Sobol :

$$\mathrm{IS}_v = \frac{\mathbb{V}\left(\phi_v(\mathbf{X}_v)\right)}{\mathbb{V}\left(\phi(\mathbf{X})\right)}$$

$$\phi(\mathbf{X}) \;=\; \phi_0 + \sum_{v \in \mathcal{P}} \phi_v(\mathbf{X}_v)$$

Objectif :

- Approcher ϕ par $f = f_0 + \sum_{v \in \mathcal{P}(d)} f_v$, telle que f_v approche ϕ_v.
- Estimer ϕ à partir $(Y_i, \mathbf{X}_i), i = 1, \ldots, n$.

Espaces d'approximation

$$\text{pour chaque } v \in \mathcal{P}, \rightarrow \ \text{RKHS } \mathcal{H}_v, \langle \cdot, \cdot \rangle_v, \ \text{noyau } k_v$$

quelques rappels :

- $\mathcal{H}_v, \langle \cdot, \cdot \rangle_v$ est un espace de Hilbert de fonctions de $L^2_{\mathbb{P}}(\mathcal{X}_v)$: $\mathbb{E}_{\mathbb{P}}(f^2(\mathbf{X}_v)) < \infty$.
- $k_v : \mathcal{X}_v \times \mathcal{X}_v \rightarrow \mathbb{R}$, symétrique, positive

$$\forall x \in \mathcal{X}_v \text{ et } \forall f_v \in \mathcal{H}_v, \ f_v(x) = \langle f_v, k_v(x, \cdot) \rangle_v$$

- $\mathcal{H}_v, \langle \cdot, \cdot \rangle_v$ est un espace de régularité : $f_v \in \mathcal{H}_v$

$$|f_v(x) - f_v(x')| \leq \|f_v\|_v \|k_v(x, \cdot) - k_v(x', \cdot)\|_v$$

- si

$$k_v(x, x') = \sum_{l=1}^{\infty} \omega_l \zeta_l(x) \zeta_l(x'), \ \text{avec } \omega_1 \geq \omega_2 \geq \ldots, \ \mathbb{E}_{\mathbb{P}}(\zeta_l \zeta_{l'}) = \delta_{l,l'}$$

alors

$$\mathcal{H}_v = \left\{ f_v = \sum_{k=1}^{\infty} a_l \zeta_l, \mathbb{E}_{\mathbb{P}}(f_v^2) = \sum_{l=1}^{\infty} a_l^2 < \infty, \|f_v\|_v^2 = \sum_{l=1}^{\infty} \frac{a_l^2}{\omega_l} < \infty \right\}$$

Espaces d'approximation de type ANOVA

$$\text{pour chaque } v \in \mathcal{P}, \rightarrow \text{ RKHS } \mathcal{H}_v, \langle \cdot, \cdot \rangle_v, \text{ noyau } k_v$$

selon une construction de type ANOVA (Durrande et al., 2013) : construire les RKHS $\mathcal{H}_v$ tels que :

$$\forall v, v' \in \mathcal{P}, v \neq v', \mathbb{E}_{\mathbb{P}} f_v(\mathbf{X}_v) = \mathbb{E}_{\mathbb{P}} f_v(\mathbf{X}_v) f_{v'}(\mathbf{X}_{v'}) = 0$$

$$\rightsquigarrow \mathcal{H} = \left\{ f : \mathcal{X} \rightarrow \mathbb{R}, f(\mathbf{X}) = f_0 + \sum_{v \in \mathcal{P}} f_v(\mathbf{X}_v), f_v \in \mathcal{H}_v \right\}$$

Les fonctions de $\mathcal{H}$ sont des candidats pour approcher ϕ :

$$f^* = \mathrm{argmin}_{f \in \mathcal{H}} \mathbb{E}_{\mathbb{P}} (\phi(\mathbf{X}) - f(\mathbf{X}))^2$$

est une approximation de la décomposition de Hoeffding-Sobol de ϕ

Estimation du méta-modèle

- Observations

$$Y_i = \phi(\mathbf{X}_i) + \varepsilon_i, i = 1, \ldots, n$$

- Espace d'approximation

$$\mathcal{H} = \left\{ f : \mathcal{X} \to \mathbb{R}, f(\mathbf{X}) = f_0 + \sum_{v \in \mathcal{P}} f_v(\mathbf{X}_v), f_v \in \mathcal{H}_v \right\}$$

$\rightsquigarrow$ procédure des moindres carrés sur les fonctions dans $\mathcal{H}$ avec pénalité de type *ridge-group-sparse*

- sélectionner les groupes v qui contribuent à bien prédire la réponse
- pour chaque groupe v sélectionné, estimer les relations entre la réponse et les variables du groupe v

Estimation du méta-modèle

basée sur une procédure *ridge-group-sparse* :

- critère des moindres carrés

$$\frac{1}{n}\sum_{i=1}^{n}\left(Y_i - f_0 - \sum_{v\in\mathcal{P}} f_v(\mathbf{X}_{iv})\right)^2 = \|\mathbf{Y} - f_0 - \sum_{v\in\mathcal{P}} \mathbf{f}_v\|_n^2$$

- pénalisé
  - une pénalité *ridge* $\rightsquigarrow$ régularité de chaque f_v à l'aide de la norme de l'espace $\mathcal{H}_v$: $\|f_v\|_v$
  - une pénalité *group-sparse* $\rightsquigarrow$ controler le nombre de groupes v nécessaires pour bien prédire ϕ :

$$\sum_v \frac{1}{\sqrt{n}}\sqrt{\sum_{i=1}^{n} f_v^2(\mathbf{X}_i)} = \sum_v \|\mathbf{f}_v\|_n$$

$$\rightsquigarrow \mathcal{L}(f) = \|\mathbf{Y} - f_0 - \sum_{v\in\mathcal{P}} \mathbf{f}_v\|_n^2 + \sum_{v\in\mathcal{P}} \mu_v \|f_v\|_v + \sum_{v\in\mathcal{P}} \gamma_v \|\mathbf{f}_v\|_n$$

Estimation du méta-modèle

$$\rightsquigarrow \mathcal{L}(f) = \|\mathbf{Y} - f_0 - \sum_{v\in\mathcal{P}} \mathbf{f}_v\|_n^2 + \sum_{v\in\mathcal{P}} \mu_v \|f_v\|_v + \sum_{v\in\mathcal{P}} \gamma_v \|\mathbf{f}_v\|_n$$

$$\widehat{f} = \operatorname{argmin}\left\{\mathcal{L}(f), f = f_0 + \sum_v f_v, f \in \mathcal{H}\right\}$$

Majoration du risque $\|\phi - \widehat{\mathbf{f}}\|_n^2$

Soit $\mathcal{F} = \{f \in \mathcal{H}, f = f_0 + \sum_v f_v,\ f_v \in \mathcal{H}_v,\ \|f_v\|_v \leq R\}$.
On note

- $\mathrm{supp}_f = \mathrm{Support}(f) : \forall v \notin \mathrm{supp}_f, f_v = 0$.
- $\nu_{n,v}$ vitesse d'estimation de f_v sur $\mathcal{H}_v$

Si

$$\gamma_v \text{ et } \sqrt{\mu_v} \geq C \max\left\{\nu_{n,v}, \sqrt{\frac{d}{n}}\right\},$$

alors, avec grande probabilité

$$\|\phi - \widehat{\mathbf{f}}\|_n^2 \leq C \inf_{f \in \mathcal{F}} \left\{\|\phi - \mathbf{f}\|_n^2 + \sum_{v \in \mathrm{supp}_f} \nu_{n,v}^2 + \frac{d|\mathrm{supp}_f|}{n}\right\}$$

Majoration du risque

Soit $\mathcal{F} = \{f \in \mathcal{H}, f = \sum_v f_v,\ f_v \in \mathcal{H}_v,\ \|f_v\|_v \leq R\}$.
avec grande probabilité

$$\|\phi - \widehat{f}\|_n^2 \leq C \inf_{f \in \mathcal{F}} \left\{ \|\phi - \mathbf{f}\|_n^2 + \sum_{v \in \mathrm{supp}_f} \nu_{n,v}^2 + \frac{d|\mathrm{supp}_f|}{n} \right\}$$

- régularité des espaces $\mathcal{H}_v$, $v \in S(f)$: $\sum_{v \in \mathrm{supp}_f} \nu_{n,v}^2$
  $\nu_{n,v}$ est lié à la vitesse d'estimation sur $\mathcal{H}_v$ (Mendelson, 2004).
  exemple : si $\omega_l \simeq (1/l)^{2\alpha}$, $\alpha > 1/2$,

$$\rightsquigarrow \nu_n \simeq n^{\frac{-\alpha}{2\alpha+1}}.$$

En particulier, si $k(x, x') = \min(x, x')$, $\nu_n = n^{-1/3}$.

Majoration du risque

Soit $\mathcal{F} = \{f \in \mathcal{H}, f = \sum_v f_v,\ f_v \in \mathcal{H}_v,\ \|f_v\|_v \leq R\}$.
avec grande probabilité

$$\|\phi - \widehat{f}\|_n^2 \leq C \inf_{f \in \mathcal{F}} \left\{ \|\phi - \mathbf{f}\|_n^2 + \sum_{v \in \mathrm{supp}_f} \nu_{n,v}^2 + \frac{d|\mathrm{supp}_f|}{n} \right\}$$

- terme de variance habituel du group-lasso :

$$\frac{d|\mathrm{supp}_f|}{n} \propto \frac{\log(|\mathcal{P}|)\,|\mathrm{supp}_f|}{n} \propto \frac{\log(2^d)\,|\mathrm{supp}_f|}{n}$$

On aimerait

- $|\mathrm{supp}_f|$ petit :
  - décomposition de Hoeffding de ϕ est « sparse »
  - bon choix des espaces $\mathcal{H}_v$ pour bien approcher ϕ
- d petit. Si d grand, restreindre aux groupes $v \in \mathcal{P}$, tels que $|v| \leq V_{\max}$,

$$\frac{\log(|\mathcal{P}|)\,|\mathrm{supp}_f|}{n} \propto \frac{\log(d)|\mathrm{supp}_f|}{n}$$

Majoration du risque

Soit $\mathcal{F} = \{f \in \mathcal{H}, f = f_0 + \sum_v f_v,\ f_v \in \mathcal{H}_v,\ \|f_v\|_v \leq R\}$.
avec grande probabilité

$$\|\phi - \widehat{f}\|_n^2 \leq C \inf_{f \in \mathcal{F}} \left\{ \|\phi - \mathbf{f}\|_n^2 + \sum_{v \in \mathrm{supp}_f} \nu_{n,v}^2 + \frac{d|\mathrm{supp}_f|}{n} \right\}$$

- $V_{\max} = 1$: modèle additif « classique »

$$f(\mathbf{X}) = f_0 + \sum_{a=1}^{d} f_a(X_a)$$

Raskutti et al. (2012), Meier et al. (2012), Koltchinsky et Yuan (2012).

Construction de $\mathcal{H}$

Durrande et al., 2013

- pour chaque coordonnée a, on choisit $(\mathcal{H}_a, k_a)$.
- les lois $\mathbb{P}_a, a = 1, \ldots, d$ sont connues.

$$a \rightsquigarrow (\mathcal{H}_a, k_a), \; \mathcal{H}_a = \mathcal{H}_{0a} \overset{\perp}{\oplus} \mathcal{H}_{1a}$$

$$\begin{aligned}
\mathcal{H}_{0a} &= \{f_a \in \mathcal{H}_a, \mathbb{E}_{\mathbb{P}_a}(f_a(X_a)) = 0\} \\
\mathcal{H}_{1a} &= \{f_a \in \mathcal{H}_a, f_a(X_a) = \text{ cste}\}
\end{aligned}$$

$$k_{0a}(X_a, X_a') = k_a(X_a, X_a') - \frac{\mathbb{E}_{U \sim \mathbb{P}_a}(k_a(X_a, U))\mathbb{E}_{U \sim \mathbb{P}_a}(k_a(X_a', U))}{\mathbb{E}_{(U,V) \sim \mathbb{P}_a \times \mathbb{P}_a} k_a(U, V)}$$

Berlinet et Thomas-Agnan (2004)

Construction de $\mathcal{H}$

$$k(\mathbf{X}, \mathbf{X}') = \prod_{a=1}^{d} \left(1 + k_{0a}(X_a, X_a')\right) = 1 + \sum_{v \in \mathcal{P}(d)} \prod_{a \in v} k_{0a}(X_a, X_a')$$

Le noyau $1 + k_{0a}(X_a, X_a')$ définit un RKHS : $\{1 \overset{\perp}{\oplus} \mathcal{H}_{0a}\}$

Le produit des noyaux définit un RKHS :

$$\mathcal{H} = \otimes_{a=1}^{d} \left(1 \overset{\perp}{\oplus} \mathcal{H}_{0a}\right) = 1 + \sum_{v \in \mathcal{P}(d)} \mathcal{H}_v$$

$$k_v(\mathbf{X}_v, \mathbf{X}_v') = \prod_{a \in v} k_{0a}(X_a, X_a')$$

- $\mathbf{X} \sim \mathbb{P} = \mathbb{P}_1 \times \ldots \times \mathbb{P}_d$, $f_v \in \mathcal{H}_v$, $f_{v'} \in \mathcal{H}_{v'}$,

  $\rightsquigarrow \mathbb{E}_{\mathbb{P}} f_v(\mathbf{X}_v) = \mathbb{E}_{\mathbb{P}} f_{v'}(\mathbf{X}_{v'}) = \mathbb{E}_{\mathbb{P}} f_v(\mathbf{X}_v) f_{v'}(\mathbf{X}_{v'}) = 0, \forall v \neq v'$

Construction de $\mathcal{H}$

$$
\begin{aligned}
k(\mathbf{X},\mathbf{X}') = \prod_{a=1}^{d}\left(1 + k_{0a}(X_a, X'_a)\right) &= 1 + \sum_{v\in\mathcal{P}(d)} \prod_{a\in v} k_{0a}(X_a, X'_a) \\
&= 1 + \sum_{v\in\mathcal{P}(d)} k_v(\mathbf{X}_v, \mathbf{X}'_v)
\end{aligned}
$$

Soit $\mathcal{H}$ le RKHS associé à k, alors $\forall f \in \mathcal{H}$,

$$
\begin{aligned}
f(\mathbf{X}) = \langle f, k(\mathbf{X},\cdot)\rangle_{\mathcal{H}} &= f_0 + \sum_{v\in\mathcal{P}} \langle f, k_v(\mathbf{X}_v,\cdot)\rangle_v \\
&= f_0 + \sum_{v\in\mathcal{P}} f_v(\mathbf{X}_v)
\end{aligned}
$$

Calcul de l'estimateur

$$\mathcal{L}(f) = \|\mathbf{Y} - f_0 - \sum_{v\in\mathcal{P}} \mathbf{f}_v\|_n^2 + \sum_{v\in\mathcal{P}} \mu_v \|f_v\|_v + \sum_{v\in\mathcal{P}} \gamma_v \|\mathbf{f}_v\|_n$$

$$\widehat{f} = \operatorname{argmin}\left\{\mathcal{L}(f), f = f_0 + \sum_v f_v, f \in \mathcal{H}\right\}$$

On se ramène à un problème paramétrique.

Propriété du RKHS $\mathcal{H}, k$:

$$\forall (\mathbf{x}_1, \ldots, \mathbf{x}_n) \in \mathcal{X}^n, (\alpha_1, \ldots, \alpha_n) \in \mathbb{R}^n,$$

$$f(\cdot) = \sum_{i=1}^n \alpha_i k(\mathbf{x}_i, \cdot) \in \mathcal{H} \text{ et}$$

$$\|f\|_{\mathcal{H}}^2 = \sum_{i,i'=1}^n \alpha_i \alpha_{i'} k(\mathbf{x}_i, \mathbf{x}_{i'})$$

Calcul de l'estimateur

En particulier la solution est de la forme

$$f = f_0 + \sum_v f_v \text{ avec } f_v(\cdot) = \sum_{i=1}^{n} \theta_{vi} k_v(\mathbf{X}_{vi}, \cdot), \boldsymbol{\theta}_v \in \mathbb{R}^n$$

Pour chaque v, soit

$$K_v \text{ matrice } n \times n, (K_v)_{i,i'} = k_v(\mathbf{X}_{vi}, \mathbf{X}_{vi'})$$

Alors

$$\widehat{f}_v(\cdot) = \widehat{f}_0 + \sum_{i=1}^{n} \widehat{\theta}_{vi} k_v(\mathbf{X}_{iv}, \cdot)$$

où les paramètres $f_0, \boldsymbol{\theta}$ minimisent

$$\|\mathbf{Y} - f_0 - \sum_v K_v \boldsymbol{\theta}_v\|_n^2 + \sum_v \gamma_v \|K_v \boldsymbol{\theta}_v\|_n + \sqrt{n} \sum_v \mu_v \|K_v^{1/2} \boldsymbol{\theta}_v\|_n$$

Calcul de l'estimateur

Algorithme de descente groupe par groupe

- CNS pour $\boldsymbol{\theta}_v = 0$. Soit

$$R_v = \mathbf{Y} - f_0 - \sum_{w \neq v} K_w \boldsymbol{\theta}_w, \text{ et } J(\mathbf{t}) = \|2R_v - n\mu_v K_v^{-1}\mathbf{t}\|^2$$

$$\boldsymbol{\theta}_v = 0 \Longleftrightarrow \min\left\{J(\mathbf{t}) \text{ sous la contrainte } \|K_v^{-1/2}\mathbf{t}\| \leq 1\right\} \leq n\gamma_v^2$$

- Si $\theta_v \neq 0$, alors

$$\theta_v = \left(\left(1 + \frac{\sqrt{n}\gamma_v}{2\|K_v\boldsymbol{\theta}_v\|}\right)K_v + \frac{n\mu_v}{2\|K_v^{1/2}\boldsymbol{\theta}_v\|}I_n\right)^{-1} R_v$$

Estimation des indices de Sobol

- Si $\mathbf{X} \sim \mathbb{P} = \mathbb{P}_1 \times \ldots \times \mathbb{P}_d$

$$\phi(\mathbf{X}) = \phi_0 + \sum_{v \in \mathcal{P}(d)} \phi_v(\mathbf{X}_v)$$

avec $E_{\mathbb{P}}\phi_v(\mathbf{X}_v) = 0$ et $E_{\mathbb{P}}\phi_v(\mathbf{X}_v)\phi_{v'}(\mathbf{X}_{v'}) = 0$ si $v \neq v'$.

- $\rightsquigarrow$ décomposition de la variance :

$$\mathbb{V}\left(\phi(\mathbf{X})\right) = \sum_{v \in \mathcal{P}(d)} \mathbb{V}\left(\phi_v(\mathbf{X}_v)\right)$$

- $\rightsquigarrow$ indices de Sobol :

$$\mathrm{IS}_v = \frac{\mathbb{V}\left(\phi_v(\mathbf{X}_v)\right)}{\mathbb{V}\left(\phi(\mathbf{X})\right)}$$

Estimation des indices de Sobol

$$\mathrm{IS}_v = \frac{\mathbb{V}\left(\phi_v(\mathbf{X}_v)\right)}{\mathbb{V}\left(\phi(\mathbf{X})\right)}$$

$\mathcal{H}, k \rightsquigarrow$ approcher ϕ par $f^* = \mathrm{argmin}_{f\in\mathcal{H}} \mathbb{E}_{\mathbb{P}}(\phi(\mathbf{X}) - f(\mathbf{X}))^2$

$$\widehat{f}(\cdot) = \widehat{f}_0 + \sum_{v\in\mathrm{supp}_{\widehat{f}}} \widehat{f}_v(\cdot) \rightsquigarrow \phi_v \leftarrow \widehat{f}_v$$

$$\begin{aligned}
\widehat{\mathrm{IS}}_v &= 0 \text{ si } v \notin \mathrm{supp}_{\widehat{f}} \\
\widehat{\mathrm{IS}}_v &= \frac{\mathbb{V}_{\mathbf{X}'\sim\mathbb{P}}\left(\widehat{f}_v(\mathbf{X}'_v)\right)}{\sum_{v\in\mathrm{supp}_{\widehat{f}}} \mathbb{V}_{\mathbf{X}'\sim\mathbb{P}}\left(\widehat{f}_v(\mathbf{X}')\right)} \text{ sinon}
\end{aligned}$$

ou bien

$$\mathbb{V}_{\mathbf{X}'\sim\mathbb{P}}\left(\widehat{f}_v(\mathbf{X}'_v)\right) \leftarrow \frac{1}{n}\sum_{i=1}^{n}\left(\widehat{f}_v(\mathbf{X}_{vi}) - \mathrm{mean}(\widehat{f}_v)\right)^2$$

Choix des paramètres (γ, μ) $(\mu_v = \mu,\ \gamma_v = \gamma)$

2 jeux de données :

- $(Y_i, \mathbf{X}_i), i = 1, \dots, n$: apprendre

$$(\gamma, \mu) \rightsquigarrow \widehat{f}_{\gamma,\mu} \text{ et } \mathrm{supp}(\widehat{f}_{\gamma,\mu}) = \widehat{S}_{\gamma,\mu}$$

- $(Y_i^t, \mathbf{X}_i^t), i = 1, \dots, n$: choisir un estimateur

2 procédures :

1. minimiser l'erreur de prédiction :

$$\left.\begin{aligned} \mathrm{EP}(\gamma, \mu) &= \textstyle\sum_{i=1}^{n} \left(Y_i^t - \widehat{f}_{\gamma,\mu}(\mathbf{X}_i^t)\right)^2 \\ (\widehat{\gamma}, \widehat{\mu}) &= \operatorname{argmin}\left\{\mathrm{EP}(\gamma, \mu), (\gamma, \mu) \in \text{ grille }\right\} \end{aligned}\right\} \rightsquigarrow \widehat{f} = \widehat{f}_{\widehat{\gamma},\widehat{\mu}}$$

Choix des paramètres (γ, μ) $(\mu_v = \mu,\ \gamma_v = \gamma)$

2 procédures :

1. minimiser l'erreur de prédiction :
2. Collection de supports

$$\left\{ \widehat{S}_{\gamma,\mu}, (\gamma,\mu) \in \text{ grille } \right\}$$

estimation ridge pour chaque support S dans la collection

$$f^{\mathrm{rdg}}_{\lambda,S} = \mathrm{argmin} \left\{ \|\mathbf{Y} - \sum_{v \in S} f_v\|_n^2 \right\} + \lambda \sum_{v \in S} \|f_v\|_v^2$$

$$\left.\begin{array}{rcl} \mathrm{EP}(\lambda, S) & = & \sum_{i=1}^{n} \left(Y_i^t - f^{\mathrm{rdg}}_{\lambda,S}(\mathbf{X}_i^t) \right)^2, \\ \widehat{\lambda} & = & \mathrm{argmin} \left\{ \mathrm{EP}(\lambda, S), \lambda \in \text{gr.}, S \in \text{coll.} \right\} \end{array}\right\} \rightsquigarrow \widehat{f} = f^{\mathrm{rdg}}_{\widehat{\lambda}(\widehat{S}),\widehat{S}}$$

Méta-modèle pour les variations du pH dans le lait

- 4 variables explicatives : $\mathbf{X}$. $n = 500$ valeurs obtenues selon un plan LHS (Latin Hypercube Sampling)
  - Volume total : lait + HCl ajouté
  - Concentration en ions H^+ ajoutés
  - Concentration de ions calcium ajoutés
  - Concentration de caséine ajoutée
- Sortie du modèle numérique $Y = m(\mathbf{X})$: pH de la solution

Modèle de régression

$$Y_i = \phi(\mathbf{X}_i) + \varepsilon_i, i = 1, \ldots, n$$

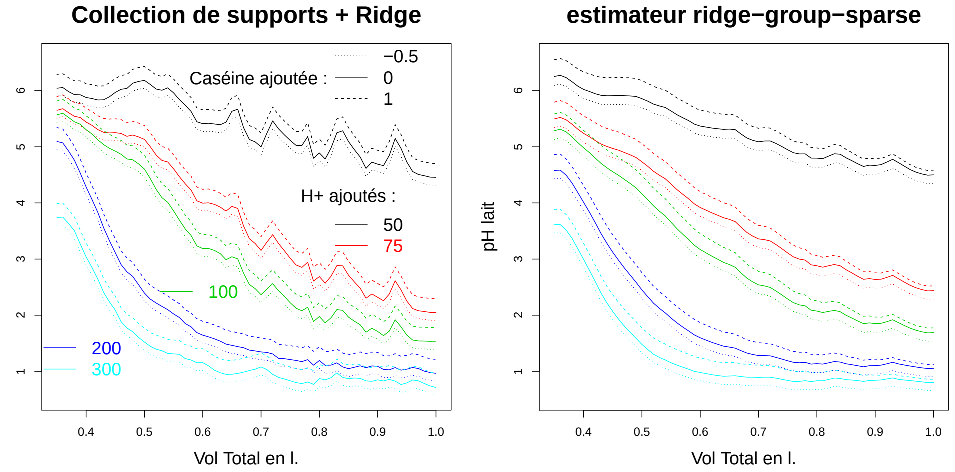
Méta-modèle pour les variations du pH dans le lait

Estimation des Indices de Sobol :

	à partir de $\widehat{f}_{\widehat{\gamma},\widehat{\mu}}$		à partir de $f^{\mathrm{rdg}}_{\widehat{\lambda},\widehat{S}}$	
groupe v	$100 \times \widehat{\mathrm{IS}}_v$	S. cum.	$100 \times \widehat{\mathrm{IS}}_v$	S. cum.
H+	75.02		72.6	
VTot	20.27	95.29	20.0	92.6
VTot, H+	3.866	99.16	6.4	99.0
Caséine	0.783	99.94	1.0	100.0
VTot, Caséine	0.055		0.0	
Ca2+	0.005		0.0	
H+, Caséine	0.002		0.0	

Méta-modèle pour les variations du pH dans le lait

pH versus (H+, VTot, Caséine)



pour terminer

- Si les $(X_1, \ldots, X_d)$ sont indépendantes de loi connue :
  - Estimation d'un méta-modèle. Majoration du risque.
  - Sélection des IS non nuls, et estimation
- Si la loi des $(X_1, \ldots, X_d)$ n'est pas connue
  - Estimation d'un méta-modèle.
    On peut toujours approcher ϕ par une fonction se décomposant selon $f^+ = f_0 + \sum_v f_v$. Mais,
    - pas de connection avec la décomposition de Hoeffding
    - pas de preuve pour la majoration du risque

Alain Berlinet and Christine Thomas-Agnan.
Reproducing kernel Hilbert spaces in probability and statistics.
Kluwer Academic, Boston, 2004.

N. Durrande, D. Ginsbourger, O. Roustant, and L. Carraro.
{ANOVA} kernels and {RKHS} of zero mean functions for model-based sensitivity analysis.
Journal of Multivariate Analysis, 115(0) :57 – 67, 2013.

Carl Holt, Douglas G. Dalgleish, and Robert Jenness.
Calculation of the ion equilibria in milk diffusate and comparison with experiment.
Analytical Biochemistry, 113(1) :154 – 163, 1981.

Vladimir Koltchinskii and Ming Yuan.
Sparsity in multiple kernel learning.
The Annals of Statistics, 38(6) :3660–3695, 12 2010.

Lukas Meier, Sara van de Geer, and Peter BÃ¼hlmann.
High-dimensional additive modeling.
The Annals of Statistics, 37(6B) :3779–3821, 12 2009.

S. Mendelson.
Geometric parameters in learning theory.
In *Geometric aspects of functional analysis*, volume 1850 of *Lecture Notes in Math.*, pages 193–235. Springer, Berlin, 2004.

Garvesh Raskutti, Martin J. Wainwright, and Bin Yu.
Minimax-optimal rates for sparse additive models over kernel classes via convex programming.
J. Mach. Learn. Res., 13 :389–427, 2012.

I. M. Sobol'.
Sensitivity estimates for nonlinear mathematical models.
Math. Model. Comput. Exp., 1(4) :407–414, 1993.

A. W. van der Vaart.
Asymptotic statistics.
Cambridge series in statistical and probabilistic mathematics. Cambridge University Press, Cambridge (UK), New York (N.Y.), 1998.