Analyse de sensibilité généralisée à base de noyaux

Application aux entrées / sorties fonctionnelles

Sébastien Da Veiga
Snecma

Séminaire MIA INRA 30/01/2015

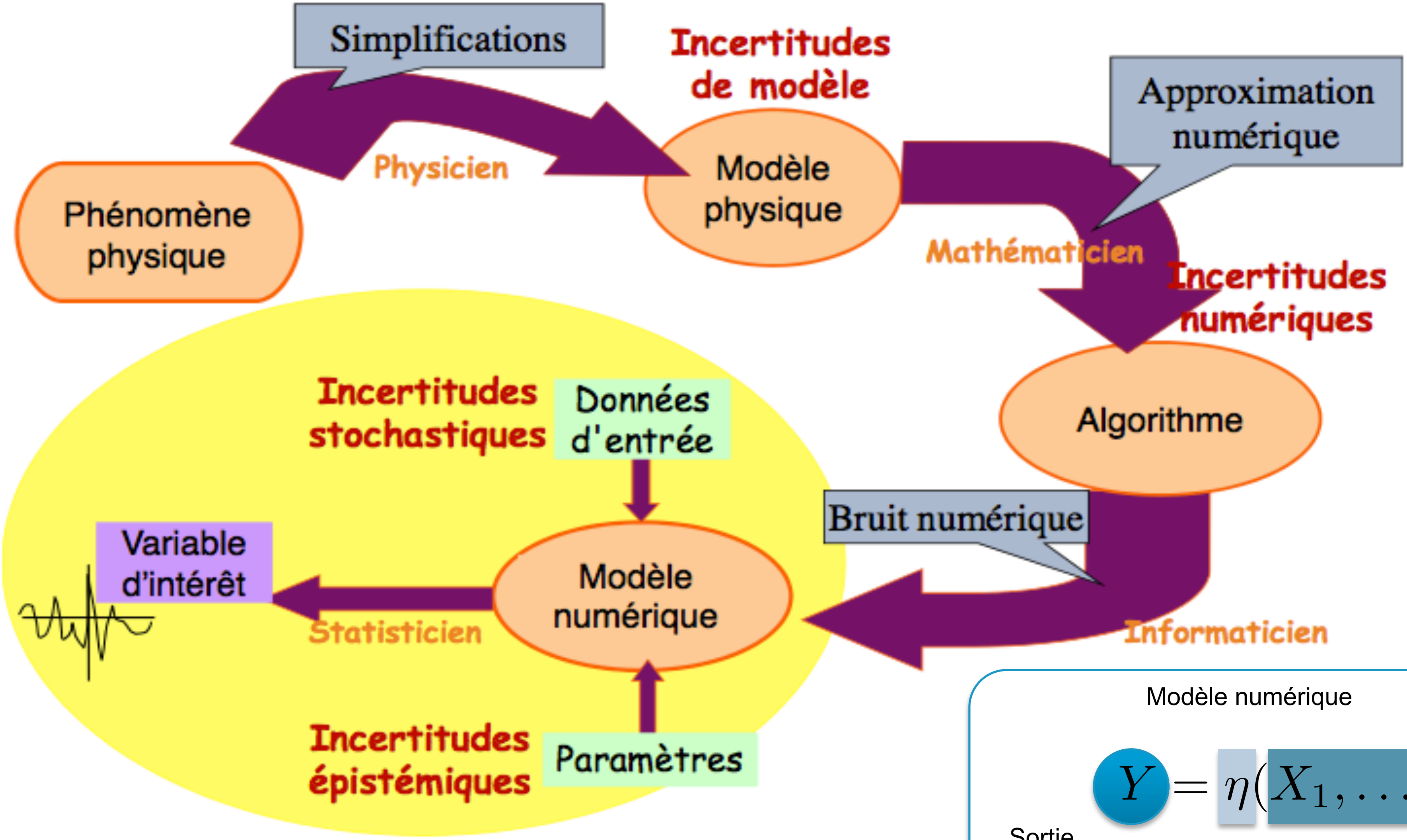
PLAN

- **Contexte et problématique**
  - Décomposition de Sobol – ANOVA fonctionnelle
- **Analyse de sensibilité généralisée**
  - Cadre général à base de densités
  - Vers de nouveaux indices de sensibilité
- **Décompositions orthogonales pour GSA généralisée**
  - Décomposition 1
  - Décomposition 2
  - Le cas des données fonctionnelles
- **Conclusions et perspectives**

/01/

CONTEXTE ET PROBLÉMATIQUE

CONTEXTE



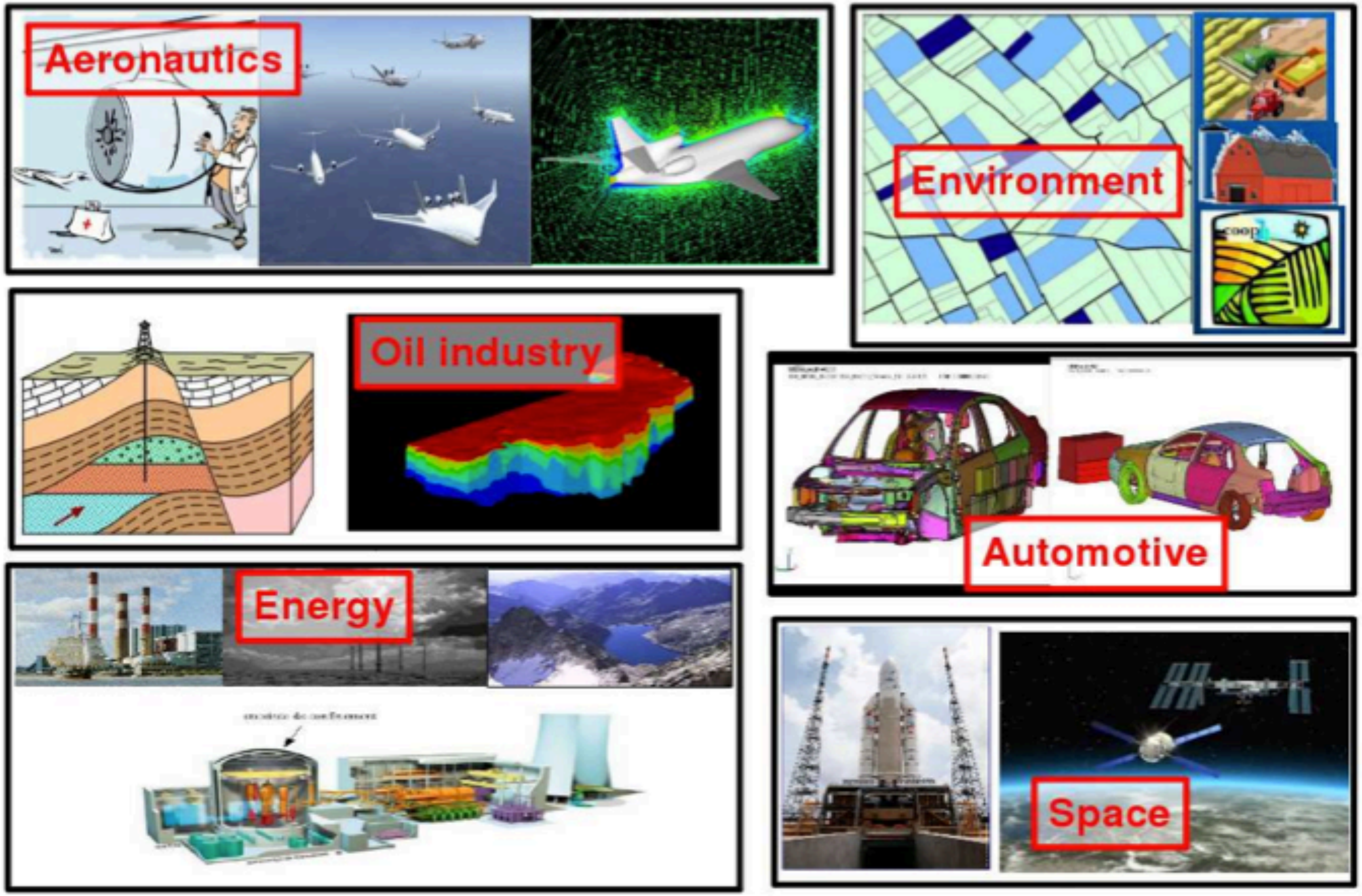
Modèle numérique

$$Y = \eta(X_1, \dots, X_d)$$

Sortie

Paramètres d'entrée

UNE PROBLÉMATIQUE MULTI-SECTORIELLE



SAFRAN Snecma

CONTEXTE

➔ Enjeux

- Comprendre l'influence des incertitudes
  - *Orienter les choix de modélisation*
  - *Prioriser les efforts de R&D*
  - *Conduire des actions supplémentaires*

- Donner du crédit à un modèle
  - *Atteindre un niveau de qualité acceptable pour son utilisation*
  - *Calibrer les paramètres du modèle*
  - *Réduire l'incertitude des sorties du modèle pour améliorer la prédiction*

- Démontrer la conformité du système avec un critère explicite ou un seuil réglementaire

- …

CONTEXTE

➔ Caractéristiques propres aux gros codes numériques

- Coût calculatoire très élevé
  - *~ 10mn – qqs heures*

- Un très grand nombre de paramètres
  - *Calibrés, réglage simulation, convergence algo résolution EDP, …*

- Des sorties complexes
  - *Courbes d'évolution, cartes ou cubes 3D de propriétés physiques, …*

- Plusieurs codes de complexité différente pour simuler le même phénomène physique
  - *Modèles analytiques, modèles 1D / 2D, choix taille maillage et discrétisation temporelle pour résolution EDP, …*

- …

SAFRAN
Snecma

CONTEXTE

➔ La démarche “standard” de traitement des incertitudes

- Comment modéliser les incertitudes sur les paramètres d’entrée ?
  - *Expériences labo, avis d’experts, calibration, …*

- Comment estimer la dispersion de la sortie d’un modèle en fonction de la dispersion des paramètres d’entrée ?
  - *MC, QMC, IS*

- Comment estimer la sensibilité de la sortie d’un code vis-à-vis d’un paramètre ou d’un groupe de paramètres d’entrée ?

CONTEXTE

➔ La démarche “standard” de traitement des incertitudes

- Comment modéliser les incertitudes sur les paramètres d’entrée ?
  - *Expériences labo, avis d’experts, calibration, …*

- Comment estimer la dispersion de la sortie d’un modèle en fonction de la dispersion des paramètres d’entrée ?
  - *MC, QMC, IS*

- Comment estimer la sensibilité de la sortie d’un code vis-à-vis d’un paramètre ou d’un groupe de paramètres d’entrée ?

CONTEXTE

➔ Analyse de sensibilité

- Objectif : identifier et hiérarchiser, parmi les paramètres d'entrée, ceux qui influencent le plus la sortie

- Pourquoi ?
  - Réduire les incertitudes sur la réponse plus efficacement en tentant de réduire l'incertitude des plus gros contributeurs
  - Améliorer la compréhension des phénomènes, guider la R&D
  - Obtenir une meilleure connaissance dans les résultats
  - Améliorer la modélisation (simplifier le modèle par exemple)

- Notations

Modèle numérique

$$\underbrace{Y}_{\text{Sortie}} = \overbrace{\eta}^{\text{Modèle numérique}}(\underbrace{X_1, \ldots, X_d}_{\text{Paramètres d'entrée}})$$

Sortie

Paramètres d'entrée

CONTEXTE

➔ Deux écoles

- Sensibilité locale : analyse le comportement de la réponse localement autour d'un point choisi (par exemple le point nominal)

$$S_i = \frac{\sigma^2_{X_i}}{\mathrm{Var}(Y)} \left(\frac{\partial \eta(X)}{\partial X_i} |_{X=X_0} \right)^2$$

  - *Très simple à mettre en oeuvre (différence finies, différenciation automatique, . . .)*
  - *MAIS approche locale, passage au global seulement sous hypothèses de linéarité du modèle*

- Sensibilité globale : faire varier l'ensemble des paramètres d'entrées du modèle dans son domaine incertain et analyser les variations des sorties

Note : liens entre les 2 approches récemment étudiés par Lamboni et al. 2013 (inégalités)

SAFRAN Snecma

CONTEXTE

➔ Analyse de sensibilité globale : 2 grandes familles

- Méthodes de criblage (screening)
  - Plans d'expériences classiques
  - Criblage
  - Méthode de Morris

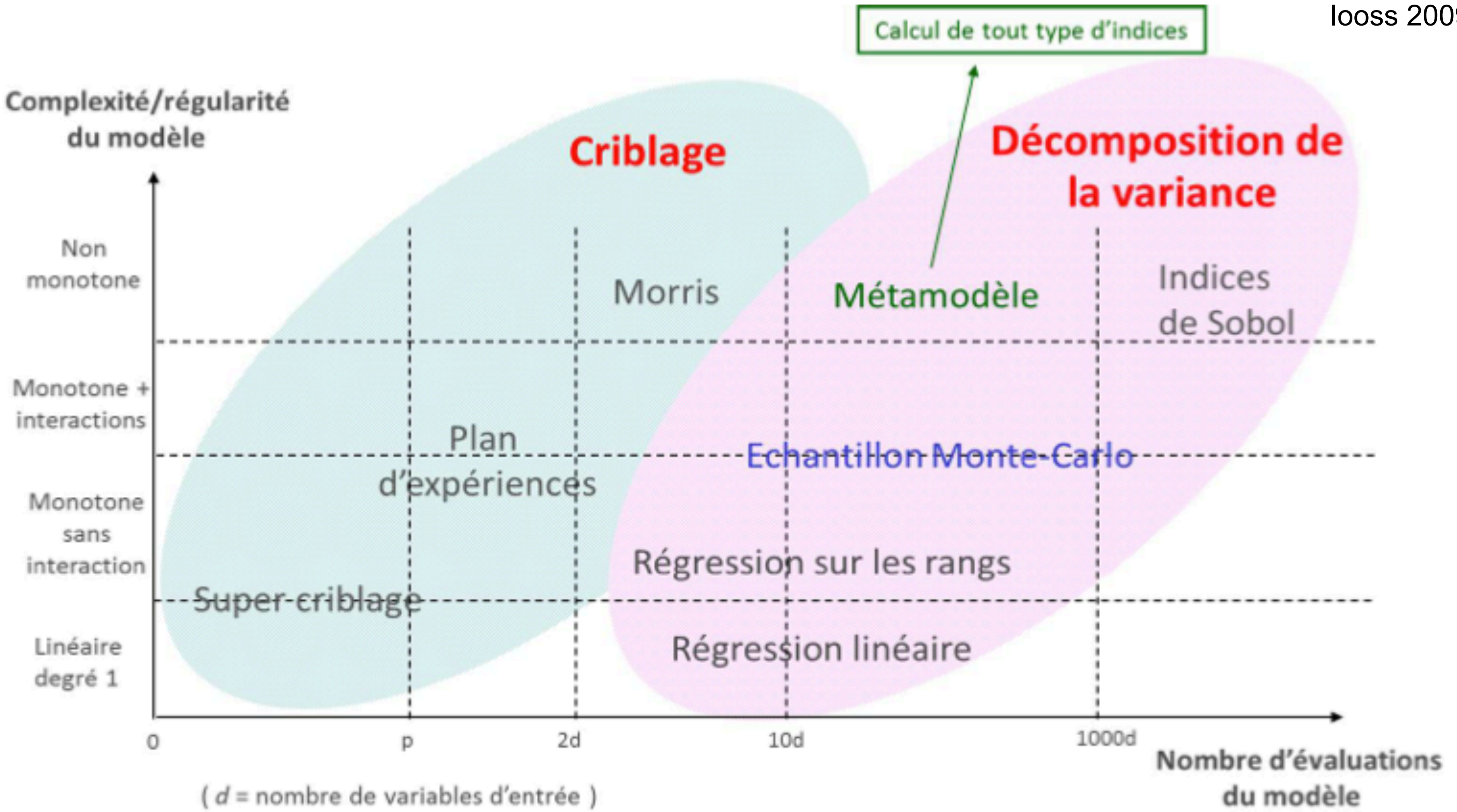
$$n \approx d/2 - 10d$$

- Méthodes quantitatives basées sur la décomposition de la variance
  - Techniques de régression linéaire (ou monotone sur rangs)
  - Indices de Sobol

$$n \approx 2d - 10^4 d$$

CONTEXTE

Iooss 2009



CONTEXTE

➔ Décomposition de la variance

- Décomposition de Sobol-Hoeffding (paramètres <u>indépendants</u>)

$$\eta(X) = \eta_0 + \sum_{i=1}^{d} \eta_i(X_i) + \sum_{1 \leq i < j \leq d} \eta_{i,j}(X_i, X_j) + \ldots + \eta_{1,\ldots,d}(X_1, \ldots, X_d)$$

  - Fonctions centrées et orthogonales
  - Expression en fonction des espérances conditionnelles :

$$\begin{aligned} \eta_0 &= \mathbb{E}(Y) \\ \eta_i(X_i) &= \mathbb{E}(Y|X_i) - \mathbb{E}(Y) \\ \eta_{i,j}(X_i, X_j) &= \mathbb{E}(Y|X_i, X_j) - \mathbb{E}(Y|X_i) - \mathbb{E}(Y|X_j) + \mathbb{E}(Y) \\ &\ldots \end{aligned}$$

SAFRAN Snecma

CONTEXTE

➔ Décomposition de la variance

- Par orthogonalité

$$\mathrm{Var}(\eta(X)) = \sum_{i=1}^{d} \mathrm{Var}(\eta_i(X_i)) + \sum_{1\leq i<j\leq d} \mathrm{Var}(\eta_{i,j}(X_i, X_j)) + \ldots + \mathrm{Var}(\eta_{1,\ldots,d}(X_1, \ldots, X_d))$$

  - Variance totale décomposée selon effets principaux, interactions, etc de chaque paramètre
  - => Indice de sensibilité pour un groupe de paramètres

$$S_I(X_I) = \frac{\mathrm{Var}(\eta_I(X_I))}{\mathrm{Var}(\eta(X))}$$

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

Effets 1er ordre

$$S_i^T(X_i) = \sum_{I \supseteq i} S_I$$

Effets totaux

SAFRAN Snecma

CONTEXTE

➔ Estimation des indices de sensibilité

- Monte-Carlo
  - Saltelli 2002, …
- Plans spéciaux
  - FAST (Cukier 1973), RBD-FAST (Mara 2009)
- Permutations (coût indépendant de la dimension)
  - Effets principaux (Mara & Joseph 2008), interactions ordre 2 (Tissot & Prieur 2014)
- Propriétés convergence, efficacité asymptotique
  - Janon et al. 2012, D. & Gamboa 2013

CONTEXTE

➔ **Résumé de la démarche pratique :**

1. Screening pour limiter le nombre de paramètres
2. Analyse quantitative pour hiérarchiser finement ceux qui restent

➔ **Mais pour des problèmes industriels sérieux**

- Coût calculatoire important du modèle numérique
  - Même les estimateurs state-of-the-art sont impossibles à utiliser
- La dimension du vecteur des entrées peut être très grande
  - Techniques de screening pas forcément robustes
- La variance de la sortie n'est pas forcément la quantité d'intérêt
  - Autres indices de sensibilité ? (e.g. pour probabilité dépassement seuil réglementaire)
- Les entrées et les sorties ne sont pas forcément des scalaires
  - Courbes d'évolution, maillages 3D, …

CONTEXTE

➔ Résumé de la démarche pratique :

1. Screening pour limiter le nombre de paramètres
2. Analyse quantitative pour hiérarchiser finement ceux qui restent

➔ Mais pour des problèmes industriels sérieux

Emulateurs

- Coût calculatoire important du modèle numérique
  - Même les estimateurs state-of-the-art sont impossibles à utiliser
- La dimension du vecteur des entrées peut être très grande
  - Techniques de screening pas forcément robustes
- La variance de la sortie n'est pas forcément la quantité d'intérêt
  - Autres indices de sensibilité ? (e.g. pour probabilité dépassement seuil réglementaire)
- Les entrées et les sorties ne sont pas forcément des scalaires
  - Courbes d'évolution, maillages 3D, …

CONTEXTE

➔ Résumé de la démarche pratique :

1. Screening pour limiter le nombre de paramètres
2. Analyse quantitative pour hiérarchiser finement ceux qui restent

➔ Mais pour des problèmes industriels sérieux

Emulateurs

- Coût calculatoire important du modèle numérique
  - Même les estimateurs state-of-the-art sont impossibles à utiliser

Feature selection

- La dimension du vecteur des entrées peut être très grande
  - Techniques de screening pas forcément robustes
- La variance de la sortie n'est pas forcément la quantité d'intérêt
  - Autres indices de sensibilité ? (e.g. pour probabilité dépassement seuil réglementaire)
- Les entrées et les sorties ne sont pas forcément des scalaires
  - Courbes d'évolution, maillages 3D, …

CONTEXTE

➔ Résumé de la démarche pratique :

1. Screening pour limiter le nombre de paramètres
2. Analyse quantitative pour hiérarchiser finement ceux qui restent

➔ Mais pour des problèmes industriels sérieux

Emulateurs

- Coût calculatoire important du modèle numérique
  - Même les estimateurs state-of-the-art sont impossibles à utiliser

Feature selection

- La dimension du vecteur des entrées peut être très grande
  - Techniques de screening pas forcément robustes

Nvx indices

- La variance de la sortie n'est pas forcément la quantité d'intérêt
  - Autres indices de sensibilité ? (e.g. pour probabilité dépassement seuil réglementaire)
- Les entrées et les sorties ne sont pas forcément des scalaires
  - Courbes d'évolution, maillages 3D, …

SAFRAN Snecma

CONTEXTE

➔ Résumé de la démarche pratique :

1. Screening pour limiter le nombre de paramètres
2. Analyse quantitative pour hiérarchiser finement ceux qui restent

➔ Mais pour des problèmes industriels sérieux

Emulateurs

- Coût calculatoire important du modèle numérique
  - Même les estimateurs state-of-the-art sont impossibles à utiliser

Feature selection

- La dimension du vecteur des entrées peut être très grande
  - Techniques de screening pas forcément robustes

Nvx indices

- La variance de la sortie n'est pas forcément la quantité d'intérêt
  - Autres indices de sensibilité ? (e.g. pour probabilité dépassement seuil réglementaire)

Distances, décompositions

- Les entrées et les sorties ne sont pas forcément des scalaires
  - Courbes d'évolution, maillages 3D, …

CONTEXTE

➔ Résumé de la démarche pratique :

1. Screening pour limiter le nombre de paramètres
2. Analyse quantitative pour hiérarchiser finement ceux qui restent

➔ Mais pour des problèmes industriels sérieux

- Coût calculatoire important du modèle numérique
  - Même les estimateurs state-of-the-art sont impossibles à utilis
- La dimension du vecteur des entrées peut être très grande
  - Techniques de screening pas forcément robustes
- La variance de la sortie n'est pas forcément la quantité d'int
  - Autres indices de sensibilité ? (e.g. pour probabilité dépassem
- Les entrées et les sorties ne sont pas forcément des scalai
  - Courbes d'évolution, maillages 3D, …

Un cadre général ?

Emulateurs

Feature selection

Nvx indices

Distances, décompositions

SAFRAN Snecma

/02/

ANALYSE DE SENSIBILITÉ GÉNÉRALISÉE

ANALYSE DE SENSIBILITÉ GÉNÉRALISÉE

➔ La variance de la sortie n'est pas forcément la quantité d'intérêt

- Très fréquemment, on est intéressé par la probabilité que la sortie dépasse un seuil (e.g. réglementaire)
- On souhaite alors identifier les paramètres d'entrée qui influencent le plus cette probabilité
  - L'analyse de sensibilité standard (à base d'indices de Sobol) ne permet pas de répondre à la question

ANALYSE DE SENSIBILITÉ GÉNÉRALISÉE

➔ La variance de la sortie n'est pas forcément la quantité d'intérêt

- Très fréquemment, on est intéressé par la probabilité que la sortie dépasse un seuil (e.g. réglementaire)
- On souhaite alors identifier les paramètres d'entrée qui influencent le plus cette probabilité
  - L'analyse de sensibilité standard (à base d'indices de Sobol) ne permet pas de répondre à la question

➔ Nécessité de définir de nouveaux indices

- Perturbation loi des entrées (Sergienko et al. 2013)
- Indices basés sur fonction de contraste (Fort et al. 2014)

$$S_i^{\psi} = \mathbb{E}\psi(Y;\theta^*) - \mathbb{E}_{(X_i,Y)}\psi(Y;\theta_i(X_i))$$

$$\theta^* = \arg\min_{\theta} \mathbb{E}\psi(Y;\theta)$$

$$\theta_i(x) = \arg\min_{\theta} \mathbb{E}(\psi(Y;\theta)|X_i = x)$$

ANALYSE DE SENSIBILITÉ GÉNÉRALISÉE

➔ **La variance de la sortie n'est qu'un résumé de la dispersion de la sortie**

- Question : est-il possible de définir un indice qui quantifie l'impact d'un paramètre d'entrée sur la loi de la sortie ?

SAFRAN Snecma

ANALYSE DE SENSIBILITÉ GÉNÉRALISÉE

➔ La variance de la sortie n'est qu'un résumé de la dispersion de la sortie

- Question : est-il possible de définir un indice qui quantifie l'impact d'un paramètre d'entrée sur la loi de la sortie ?

- Indices « distributionnels »

$$S_i^{TV} = \int |p_Y(y) - p_{Y|X_i=x}(y)| p_{X_i}(x) dx dy$$

Borgonovo 2007

$$S_i^{KL} = \int p_{Y|X_i=x}(y) \ln \left(\frac{p_{Y|X_i=x}(y)}{p_Y(y)} \right) p_{X_i}(x) dx dy$$

Kraskov et al. 2001

ANALYSE DE SENSIBILITÉ GÉNÉRALISÉE

➔ La variance de la sortie n'est qu'un résumé de la dispersion de la sortie

- Question : est-il possible de définir un indice qui quantifie l'impact d'un paramètre d'entrée sur la loi de la sortie ?

- Indices « distributionnels »

$$S_i^{TV} = \int |p_Y(y) - p_{Y|X_i=x}(y)| p_{X_i}(x) dx dy$$

Borgonovo 2007

$$S_i^{KL} = \int p_{Y|X_i=x}(y) \ln\left(\frac{p_{Y|X_i=x}(y)}{p_Y(y)}\right) p_{X_i}(x) dx dy$$

Kraskov et al. 2001

- On reconnaît facilement la distance TV et l'information mutuelle (j'y reviendrai)

ANALYSE DE SENSIBILITÉ GÉNÉRALISÉE

➔ Cadre général pour les indices distributionnels

- En toute généralité, l'impact d'un paramètre d'entrée peut être défini par l'intermédiaire d'une mesure de dissimilarité entre lois de probabilité

$$S_i = \mathbb{E}_{X_i}\left(d(P_Y, P_{Y|X_i})\right)$$

Baucells and Borgonovo 2013
D. 2014

  - Si la loi de la sortie et la loi conditionnelle sont « proches », le paramètre n'a pas d'influence

SAFRAN Snecma

ANALYSE DE SENSIBILITÉ GÉNÉRALISÉE

➔ Cadre général pour les indices distributionnels

- En toute généralité, l'impact d'un paramètre d'entrée peut être défini par l'intermédiaire d'une mesure de dissimilarité entre lois de probabilité

$$S_i = \mathbb{E}_{X_i}\left(d(P_Y, P_{Y|X_i})\right)$$

Baucells and Borgonovo 2013
D. 2014

  - Si la loi de la sortie et la loi conditionnelle sont « proches », le paramètre n'a pas d'influence

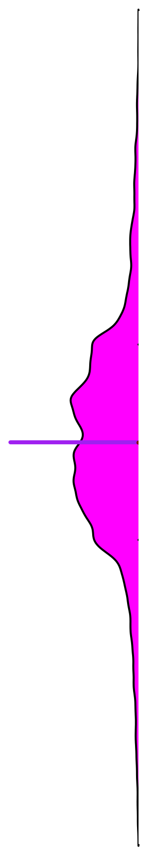
- Un exemple qui va nous servir tout au long de cet exposé

$$Y = \sin(X_1) + 5\sin^2(X_2) + 0.1 X_3^4 \sin(X_1)$$

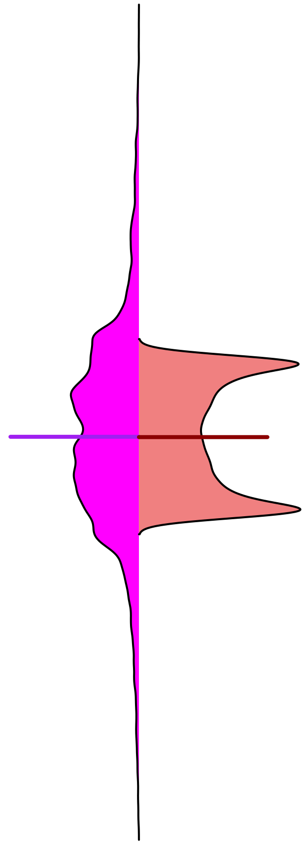
$$X_1, X_2, X_3, X_4 \sim U\left(-\pi, \pi\right)$$

Fonction d'Ishigami avec variable dummy

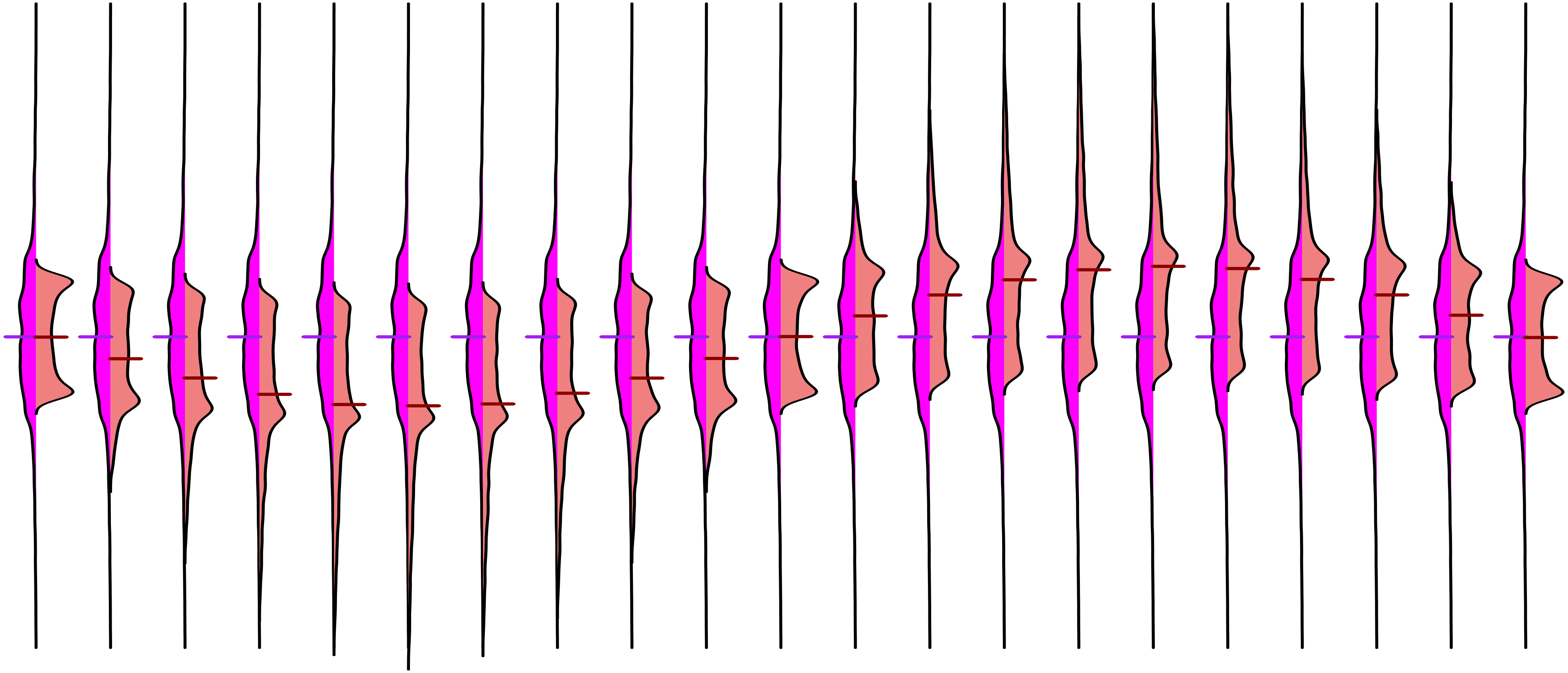
ANALYSE DE SENSIBILITÉ GÉNÉRALISÉE

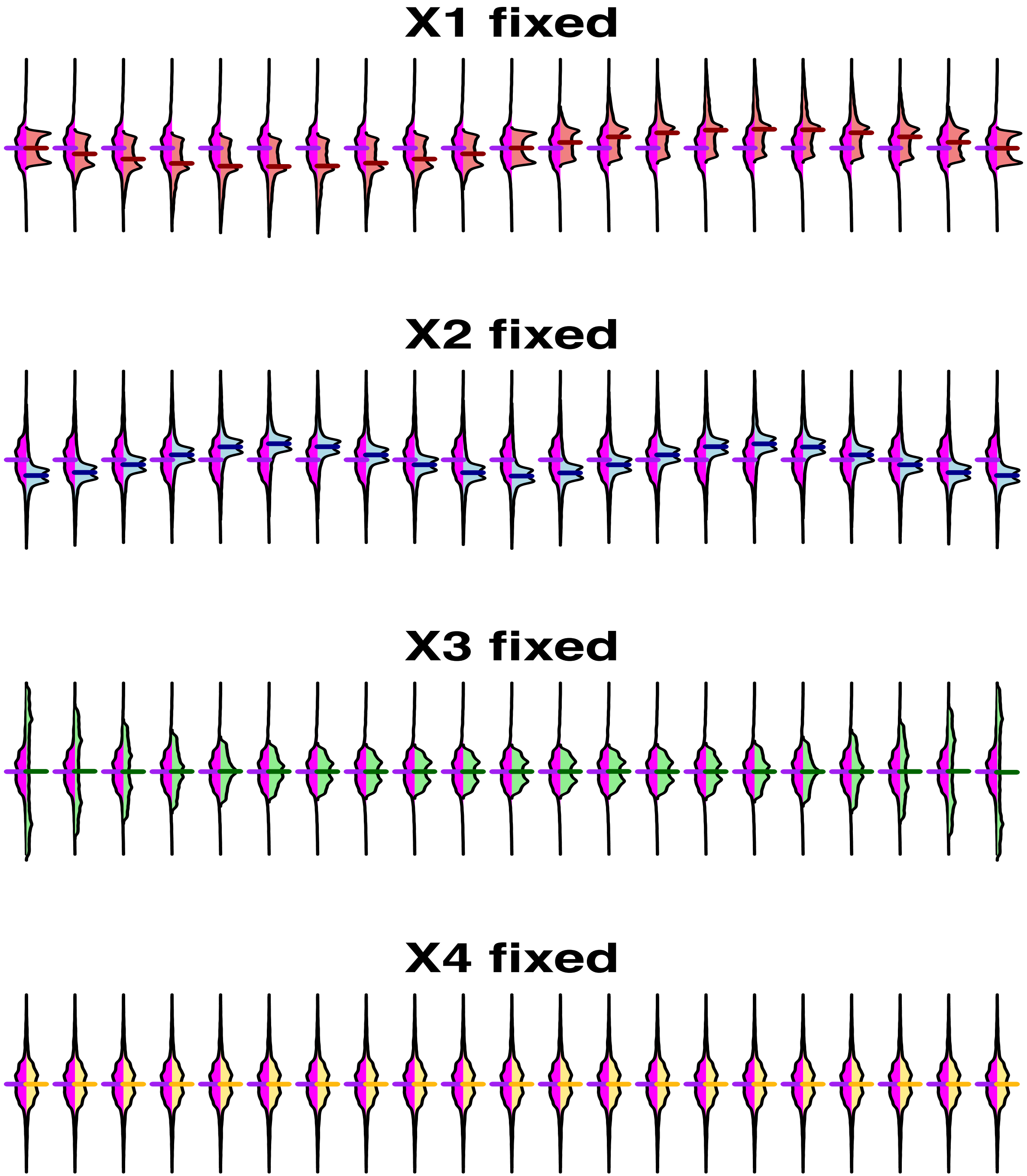


ANALYSE DE SENSIBILITÉ GÉNÉRALISÉE



ANALYSE DE SENSIBILITÉ GÉNÉRALISÉE





Analyse visuelle ?

ANALYSE DE SENSIBILITÉ GÉNÉRALISÉE

➔ Comment comparer des lois de probabilité ?

- Le plus « basique »

ANALYSE DE SENSIBILITÉ GÉNÉRALISÉE

➔ Comment comparer des lois de probabilité ?

- Le plus « basique »
  - Comparer leur moyenne

$$d(P_Y, P_{Y|X_i}) = (\mathbb{E}(Y) - \mathbb{E}(Y|X_i))^2$$

ANALYSE DE SENSIBILITÉ GÉNÉRALISÉE

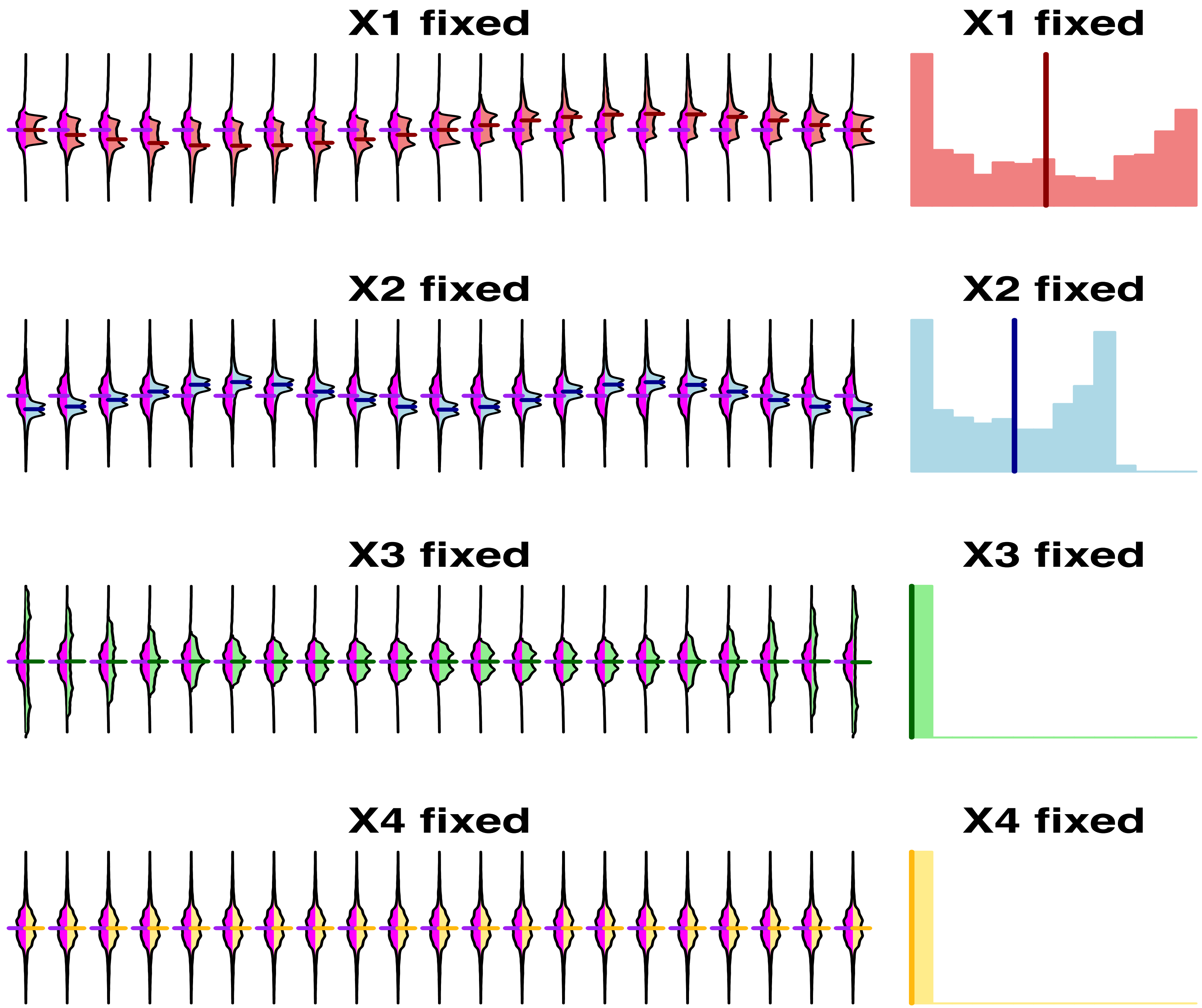
➔ Comment comparer des lois de probabilité ?

- Le plus « basique »
  - Comparer leur moyenne

$$d(P_Y, P_{Y|X_i}) = (\mathbb{E}(Y) - \mathbb{E}(Y|X_i))^2$$

➔ Sobol !

SAFRAN Snecma


X1 fixed
X1 fixed
X2 fixed
X2 fixed
X3 fixed
X3 fixed
X4 fixed
X4 fixed

ANALYSE DE SENSIBILITÉ GÉNÉRALISÉE

➔ Comment comparer des lois de probabilité ?

- Le plus « basique »
  - Comparer leur moyenne

$$d(P_Y, P_{Y|X_i}) = (\mathbb{E}(Y) - \mathbb{E}(Y|X_i))^2$$

➔ Sobol !

- La famille des f-divergences

$$d_f(P_Y || P_{Y|X_i}) = \int f\left(\frac{p_Y(y)}{p_{Y|X_i}(y)}\right) p_{Y|X_i}(y) dy$$

SAFRAN Snecma

ANALYSE DE SENSIBILITÉ GÉNÉRALISÉE

➔ Comment comparer des lois de probabilité ?

- Le plus « basique »
  - Comparer leur moyenne

$$d(P_Y, P_{Y|X_i}) = (\mathbb{E}(Y) - \mathbb{E}(Y|X_i))^2$$

➔ Sobol !

- La famille des f-divergences

$$d_f(P_Y||P_{Y|X_i}) = \int f\left(\frac{p_Y(y)}{p_{Y|X_i}(y)}\right) p_{Y|X_i}(y)dy$$

$$S_i^f = \int f\left(\frac{p_Y(y)p_{X_i}(x)}{p_{X_i,Y}(x,y)}\right) p_{X_i,Y}(x,y)dxdy$$

D. 2014

SAFRAN Snecma

ANALYSE DE SENSIBILITÉ GÉNÉRALISÉE

➔ Comment comparer des lois de probabilité ?

- Le plus « basique »
  - Comparer leur moyenne

$$d(P_Y, P_{Y|X_i}) = (\mathbb{E}(Y) - \mathbb{E}(Y|X_i))^2$$

➔ Sobol !

- La famille des f-divergences

$$d_f(P_Y || P_{Y|X_i}) = \int f\left(\frac{p_Y(y)}{p_{Y|X_i}(y)}\right) p_{Y|X_i}(y)dy$$

  - Inclut comme cas particuliers d'indices de sensibilité la distance TV et l'information mutuelle

$$S_i^{TV} = \int |p_Y(y) - p_{Y|X_i=x}(y)| p_{X_i}(x)dxdy$$

Borgonovo 2007

$$S_i^{KL} = \int p_{Y|X_i=x}(y) \ln\left(\frac{p_{Y|X_i=x}(y)}{p_Y(y)}\right) p_{X_i}(x)dxdy$$

Kraskov et al. 2001

ANALYSE DE SENSIBILITÉ GÉNÉRALISÉE

➔ Avantages de ces indices

- Prise en compte de l'effet général d'un paramètre sur la loi de la sortie et pas seulement en moyenne
- A base de densités
  - Beaucoup de codes/routines existantes pour l'estimation
  - Plusieurs distances possibles (sans coût de calcul supplémentaire)

ANALYSE DE SENSIBILITÉ GÉNÉRALISÉE

➔ Avantages de ces indices

- Prise en compte de l'effet général d'un paramètre sur la loi de la sortie et pas seulement en moyenne
- A base de densités
  - Beaucoup de codes/routines existantes pour l'estimation
  - Plusieurs distances possibles (sans coût de calcul supplémentaire)

➔ Un problème ?

- Penser par exemple au cas où la sortie est multi-dimensionnelle

ANALYSE DE SENSIBILITÉ GÉNÉRALISÉE

➔ Avantages de ces indices

- Prise en compte de l'effet général d'un paramètre sur la loi de la sortie et pas seulement en moyenne
- A base de densités
  - Beaucoup de codes/routines existantes pour l'estimation
  - Plusieurs distances possibles (sans coût de calcul supplémentaire)

➔ Un problème ?

- Penser par exemple au cas où la sortie est multi-dimensionnelle
- L'estimation de densité en grande dimension est très difficile
- Biais des estimateurs

ANALYSE DE SENSIBILITÉ GÉNÉRALISÉE

➔ Avantages de ces indices

- Prise en compte de l'effet général d'un paramètre sur la loi de la sortie et pas seulement en moyenne
- A base de densités
  - Beaucoup de codes/routines existantes pour l'estimation
  - Plusieurs distances possibles (sans coût de calcul supplémentaire)

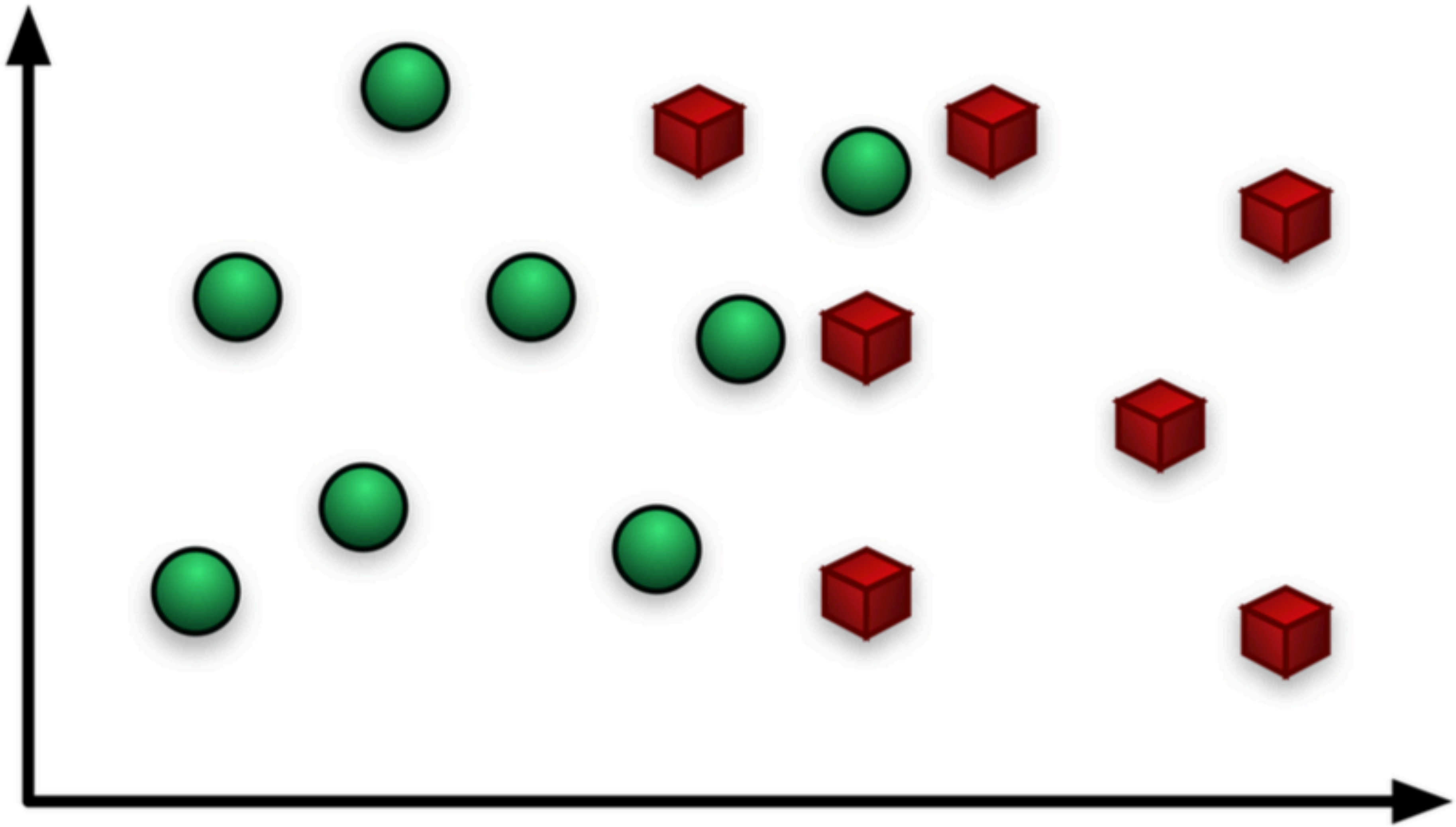
➔ Un problème ?

- Penser par exemple au cas où la sortie est multi-dimensionnelle
- L'estimation de densité en grande dimension est très difficile
- Biais des estimateurs

➔ Travaux récents pour comparer des distributions

- La moyenne n'est peut-être pas si bête …
- … et son estimation est beaucoup plus aisée

ANALYSE DE SENSIBILITÉ GÉNÉRALISÉE



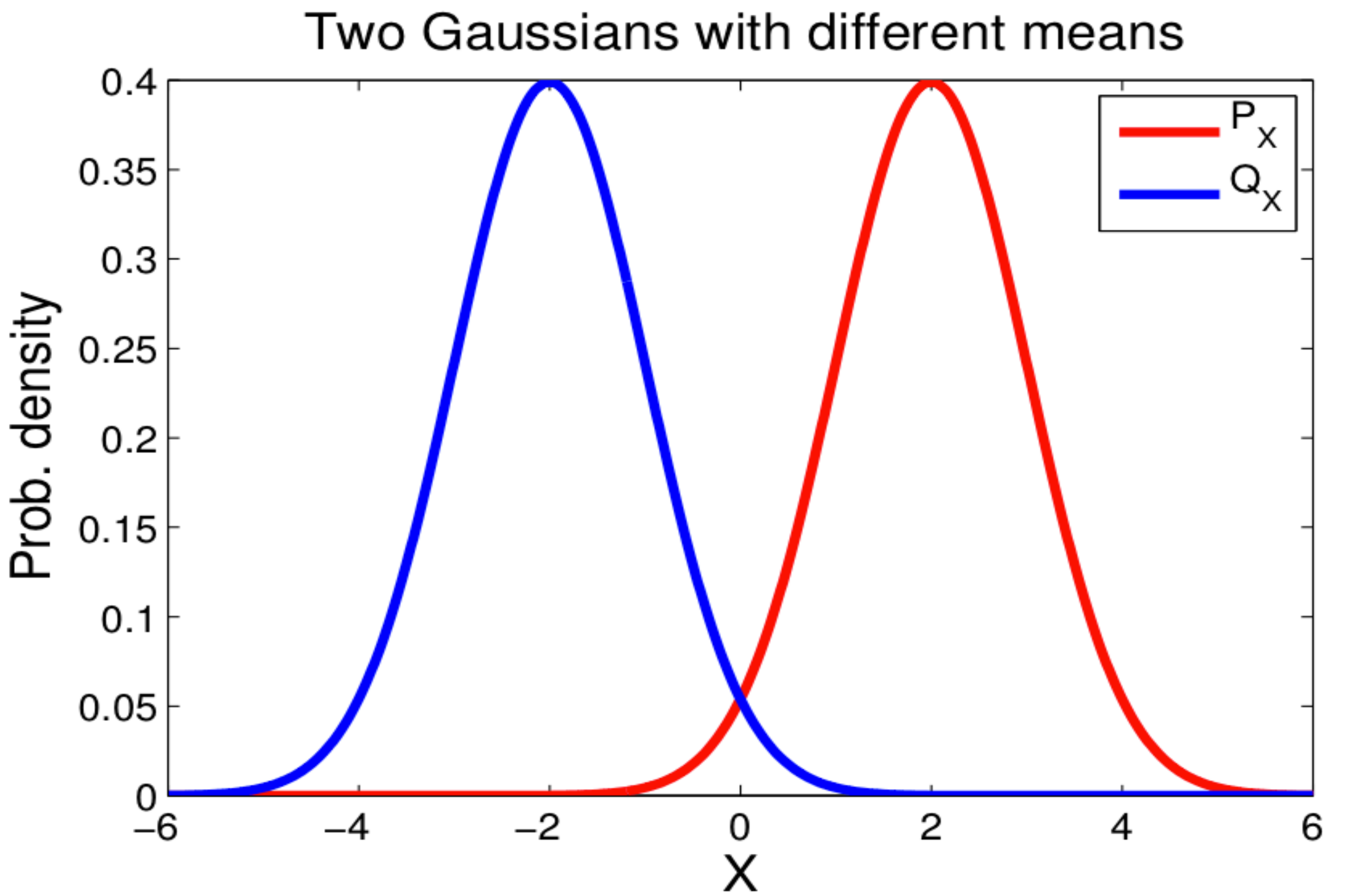
Smola 2006

ANALYSE DE SENSIBILITÉ GÉNÉRALISÉE

➔ Idée

- Pas besoin d'estimation de densité si le problème est vraiment simple
- Dans l'exemple de Smola, il suffit de comparer les moyennes

ANALYSE DE SENSIBILITÉ GÉNÉRALISÉE



Gretton 2012

ANALYSE DE SENSIBILITÉ GÉNÉRALISÉE

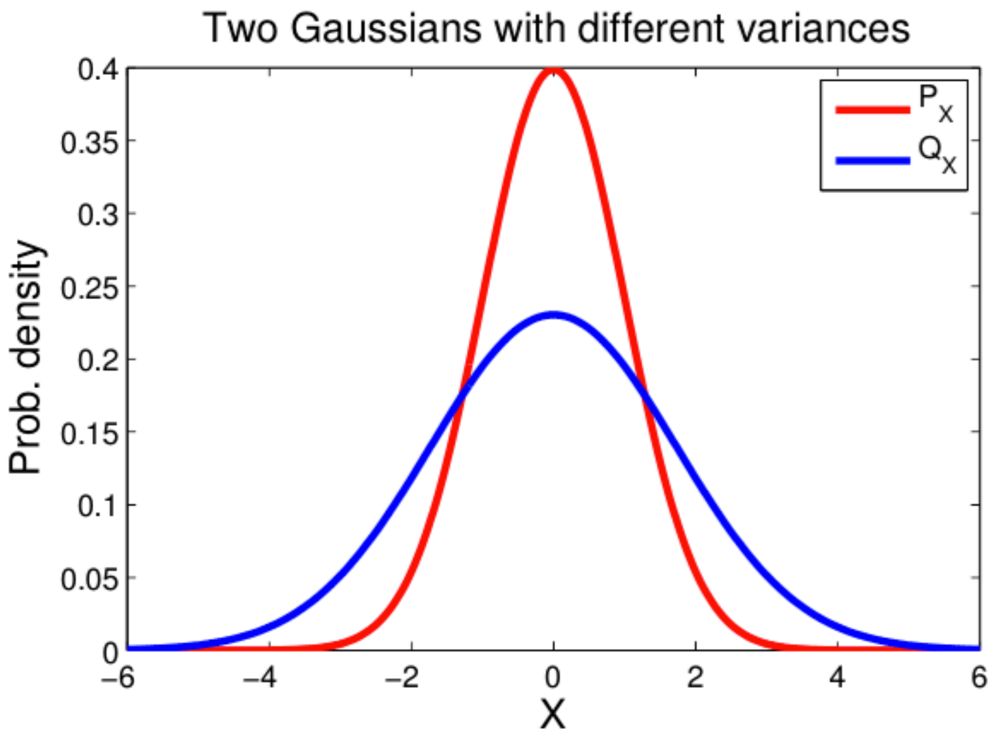
➔ Idée

- Pas besoin d'estimation de densité si le problème est vraiment simple
- Dans l'exemple de Smola, il suffit de comparer les moyennes

➔ De manière générale

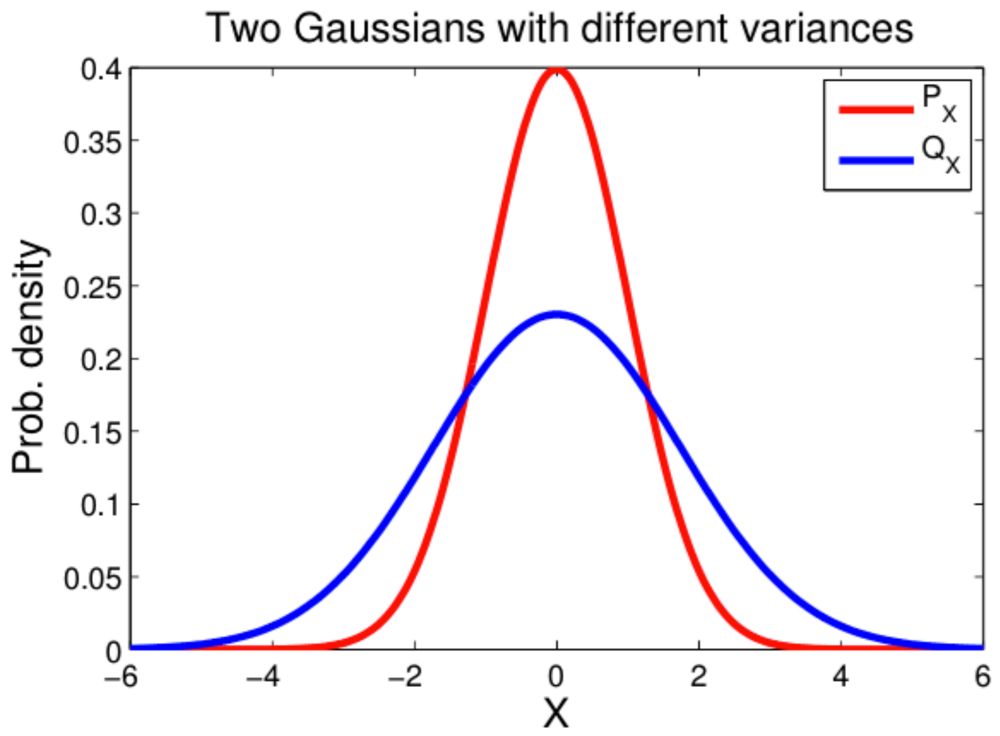
- Regarder la différence entre les moyennes de caractéristiques (« features) des variables aléatoires

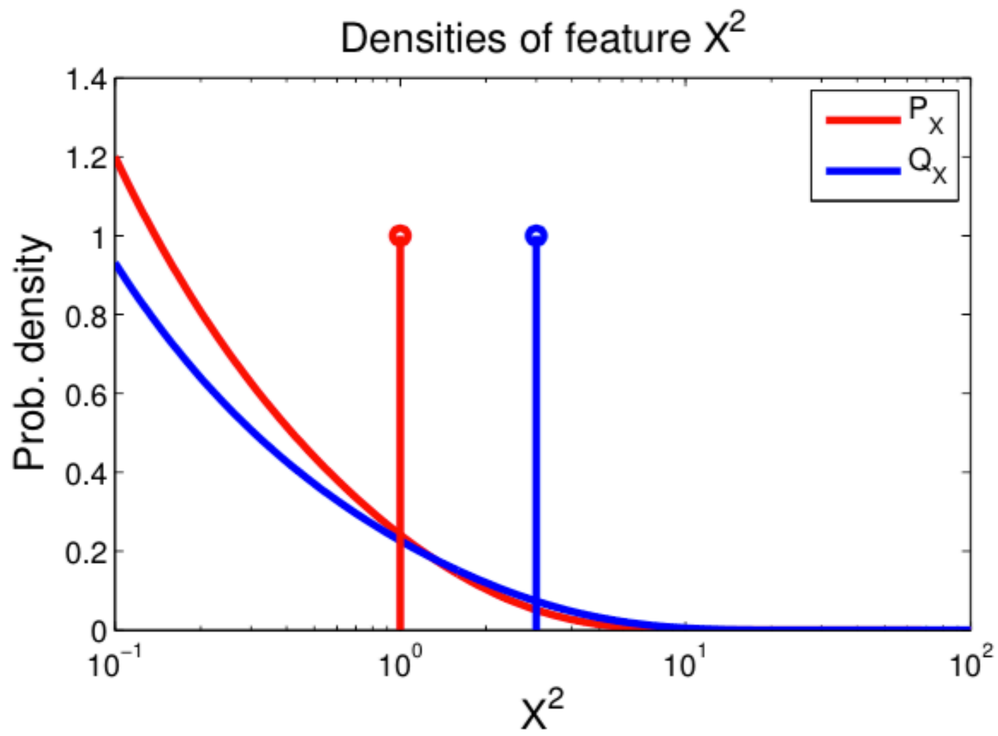
ANALYSE DE SENSIBILITÉ GÉNÉRALISÉE



Gretton 2012

ANALYSE DE SENSIBILITÉ GÉNÉRALISÉE





Gretton 2012

ANALYSE DE SENSIBILITÉ GÉNÉRALISÉE

➔ Idée

- Pas besoin d'estimation de densité si le problème est vraiment simple
- Dans l'exemple de Smola, il suffit de comparer les moyennes

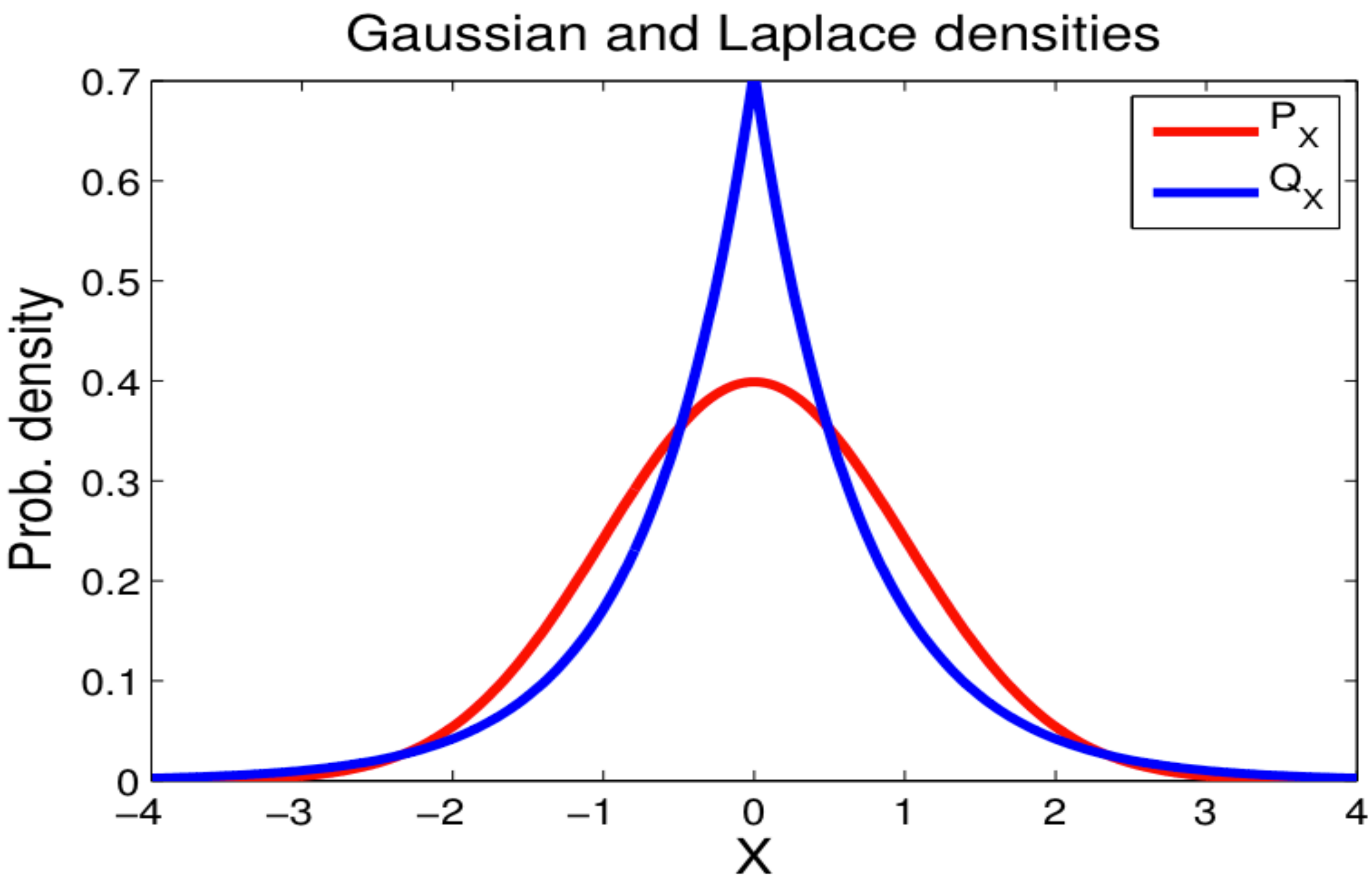
➔ De manière générale

- Regarder la différence entre les moyennes de caractéristiques (« features) des variables aléatoires
- Dans le 2ème exemple de Gretton, il suffit de considérer des features d'ordre 2 et de comparer leurs moyennes

$$\varphi_x = x^2$$

SAFRAN Snecma

ANALYSE DE SENSIBILITÉ GÉNÉRALISÉE



Gretton 2012

ANALYSE DE SENSIBILITÉ GÉNÉRALISÉE

➔ Idée

- Pas besoin d'estimation de densité si le problème est vraiment simple
- Dans l'exemple de Smola, il suffit de comparer les moyennes

➔ De manière générale

- Regarder la différence entre les moyennes de caractéristiques (« features) des variables aléatoires
- Dans le 2ème exemple de Gretton, il suffit de considérer des features d'ordre 2 et de comparer leurs moyennes

$$\varphi_x = x^2$$

- Dans le 3ème exemple de Gretton, il faut étudier la différence des moyennes de features d'ordre plus élevé

ANALYSE DE SENSIBILITÉ GÉNÉRALISÉE

➔ Théorie générale : Maximum Mean Discrepancy

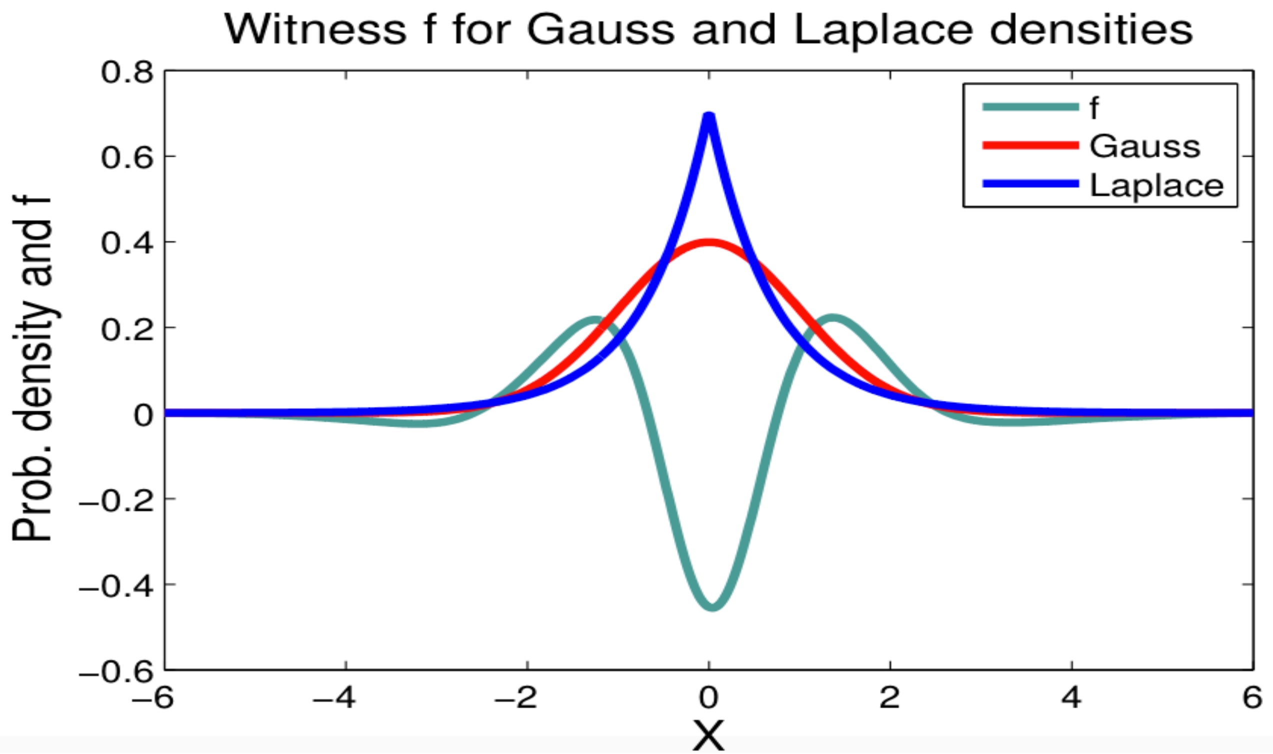
$$\mathrm{MMD}(P, Q; F) := \sup_{f \in F} \left[\mathbb{E}_P f(x) - \mathbb{E}_Q f(x) \right]$$

Gretton 2012

ANALYSE DE SENSIBILITÉ GÉNÉRALISÉE

➔ Théorie générale : Maximum Mean Discrepancy

$$\mathrm{MMD}(P, Q; F) := \sup_{f \in F} \left[\mathbb{E}_P f(x) - \mathbb{E}_Q f(x) \right]$$



Gretton 2012

ANALYSE DE SENSIBILITÉ GÉNÉRALISÉE

➔ Théorie générale : Maximum Mean Discrepancy

$$\mathrm{MMD}(P, Q; F) := \sup_{f \in F} \left[\mathbb{E}_P f(x) - \mathbb{E}_Q f(x) \right]$$

➔ Résultats classiques : distance nulle ssi les lois sont égales quand

- F = fonctions continues bornées (métrique de Dudley)
- F = fonctions de variations bornées à 1 (métrique de Kolmogorov)
- F = fonctions bornées Lipschitz (« Earth mover's distance »)

ANALYSE DE SENSIBILITÉ GÉNÉRALISÉE

➔ Théorie générale : Maximum Mean Discrepancy

$$\mathrm{MMD}(P, Q; F) := \sup_{f \in F} \left[\mathbb{E}_P f(x) - \mathbb{E}_Q f(x) \right]$$

➔ Résultats classiques : distance nulle ssi les lois sont égales quand

- F = fonctions continues bornées (métrique de Dudley)
- F = fonctions de variations bornées à 1 (métrique de Kolmogorov)
- F = fonctions bornées Lipschitz (« Earth mover's distance »)

➔ Résultat plus récent

- F = boule unité dans un RKHS caractéristique

ANALYSE DE SENSIBILITÉ GÉNÉRALISÉE

➔ Fonctions dans un RKHS

$$\mathcal{F} \text{ RKHS de } \mathcal{X} \text{ dans } \mathbb{R} \text{ avec noyau défini positif } k(x, x')$$

$$\mathcal{F} = \overline{\mathrm{span}\left\{k(x, \cdot) | x \in \mathcal{X}\right\}}$$

$$f(x) = \sum_{i=1}^{m} \alpha_i k(x_i, x) \text{ pour } m \in \mathbb{N}, \alpha_i \in \mathbb{R}, x_i \in \mathcal{X} \text{ arbitraires}$$

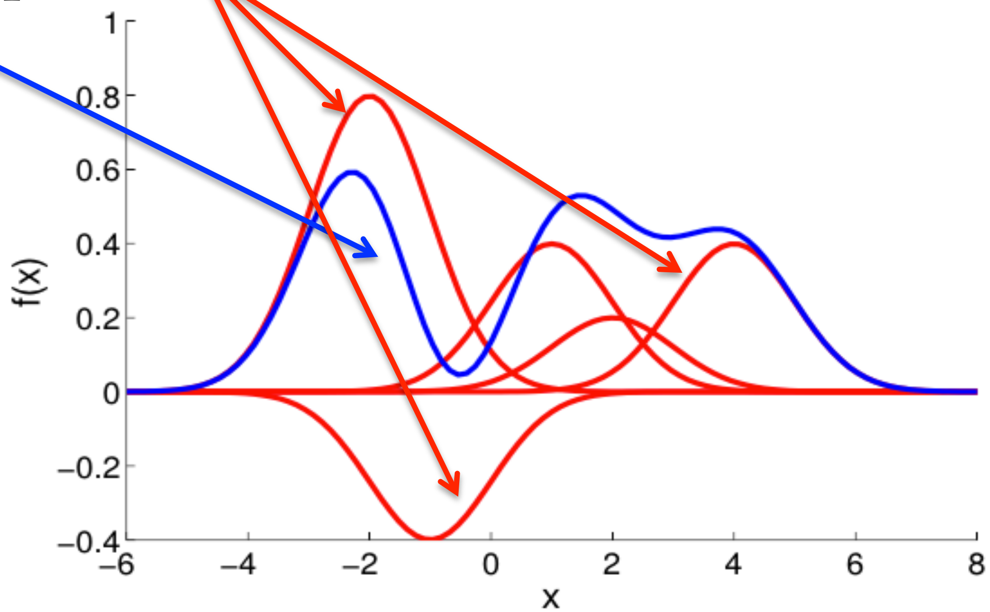
ANALYSE DE SENSIBILITÉ GÉNÉRALISÉE

➔ Fonctions dans un RKHS

$$\mathcal{F} \text{ RKHS de } \mathcal{X} \text{ dans } \mathbb{R} \text{ avec noyau défini positif } k(x, x')$$

$$\mathcal{F} = \overline{\text{span}\left\{k(x, \cdot) | x \in \mathcal{X}\right\}}$$

$$\boxed{f(x)} = \sum_{i=1}^{m} \alpha_i \boxed{k(x_i, x)} \text{ pour } m \in \mathbb{N}, \alpha_i \in \mathbb{R}, x_i \in \mathcal{X} \text{ arbitraires}$$



ANALYSE DE SENSIBILITÉ GÉNÉRALISÉE

➔ **Lien avec les features vues précédemment ?**

ANALYSE DE SENSIBILITÉ GÉNÉRALISÉE

➔ **Lien avec les features vues précédemment ?**

$$x \in \mathbb{R}^2, \text{ feature } \varphi_x$$

$$\varphi_x^{(p)} = \left[x_1^2 \quad x_2^2 \quad x_1 x_2 \sqrt{2} \right]$$

$$\varphi_x^{(g)} = \exp\left(-\lambda \|x - \cdot\|^2 \right)$$

ANALYSE DE SENSIBILITÉ GÉNÉRALISÉE

➔ Lien avec les features vues précédemment ?

$$x \in \mathbb{R}^2, \text{ feature } \varphi_x$$

$$\varphi_x^{(p)} = \begin{bmatrix} x_1^2 & x_2^2 & x_1 x_2 \sqrt{2} \end{bmatrix} \qquad \varphi_x^{(g)} = \exp\left(-\lambda \|x - \cdot\|^2\right)$$

- Produit scalaire entre fonctions features

$$\left\langle \varphi_x^{(p)}, \varphi_{x'}^{(p)} \right\rangle_{\mathcal{F}} = \langle x, x' \rangle^2 \qquad \left\langle \varphi_x^{(g)}, \varphi_{x'}^{(g)} \right\rangle_{\mathcal{F}} = \exp\left(-\lambda \|x - x'\|^2\right)$$

ANALYSE DE SENSIBILITÉ GÉNÉRALISÉE

➔ Lien avec les features vues précédemment ?

$$x \in \mathbb{R}^2, \ \text{feature } \varphi_x$$

$$\varphi_x^{(p)} = \begin{bmatrix} x_1^2 & x_2^2 & x_1 x_2 \sqrt{2} \end{bmatrix} \qquad \varphi_x^{(g)} = \exp\left(-\lambda \|x - \cdot\|^2\right)$$

- Produit scalaire entre fonctions features

$$\left\langle \varphi_x^{(p)}, \varphi_{x'}^{(p)} \right\rangle_{\mathcal{F}} = \langle x, x' \rangle^2 \qquad \left\langle \varphi_x^{(g)}, \varphi_{x'}^{(g)} \right\rangle_{\mathcal{F}} = \exp\left(-\lambda \|x - x'\|^2\right)$$

- De manière générale

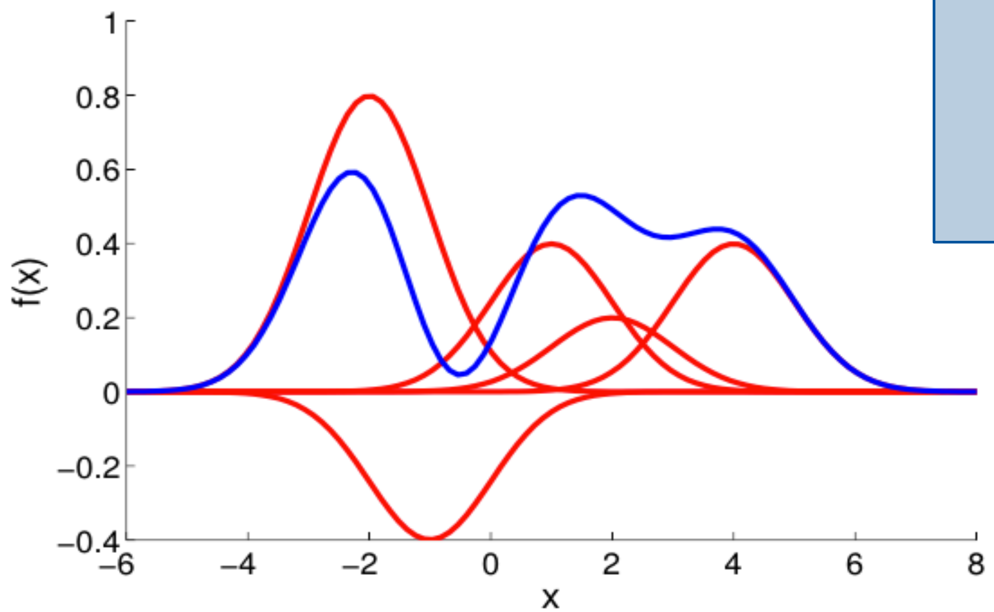
$$\boxed{\langle \varphi_x, \varphi_{x'} \rangle_{\mathcal{F}} = k(x, x')}$$

Les noyaux sont des produits scalaires de fonctions features

ANALYSE DE SENSIBILITÉ GÉNÉRALISÉE

➔ Exemple

$$f(x) = \sum_{i=1}^{m} \alpha_i k(x_i, x) = \sum_{i=1}^{m} \alpha_i \langle \varphi_{x_i}, \varphi_x \rangle_{\mathcal{F}} = \langle f, \varphi_x \rangle_{\mathcal{F}}$$



$$f = \sum_{i=1}^{m} \alpha_i \varphi_{x_i}$$

SAFRAN Snecma

ANALYSE DE SENSIBILITÉ GÉNÉRALISÉE

➔ Finalement, Maximum Mean Discrepancy dans le cadre RKHS

$$\mathrm{MMD}^2(P,Q;F) \quad = \left(\sup_{f \in F} \left[\mathbb{E}_P f(x) - \mathbb{E}_Q f(x) \right] \right)^2$$

ANALYSE DE SENSIBILITÉ GÉNÉRALISÉE

➔ **Finalement, Maximum Mean Discrepancy dans le cadre RKHS**

$$\begin{aligned}\mathrm{MMD}^2(P,Q;F) &= \left(\sup_{f\in F}\left[\mathbb{E}_P f(x) - \mathbb{E}_Q f(x)\right]\right)^2 \\ &= \left(\sup_{f\in F}\langle f, \mu_P - \mu_Q\rangle_{\mathcal{F}}\right)^2\end{aligned}$$

$$\begin{aligned}\mathbb{E}_P f(x) &= \mathbb{E}_P\left[\langle \varphi_x, f\rangle_{\mathcal{F}}\right] \\ &:= \langle \mu_P, f\rangle_{\mathcal{F}}\end{aligned}$$

$$\mu_P = \mathbb{E}_P \varphi_x = \int k(x,\cdot)\,dP(x)$$

ANALYSE DE SENSIBILITÉ GÉNÉRALISÉE

➔ Finalement, Maximum Mean Discrepancy dans le cadre RKHS

$$
\begin{aligned}
\mathrm{MMD}^2(P,Q;F) &= \left(\sup_{f\in F}\left[\mathbb{E}_P f(x) - \mathbb{E}_Q f(x)\right]\right)^2 \\
&= \left(\sup_{f\in F}\langle f, \mu_P - \mu_Q\rangle_{\mathcal{F}}\right)^2 \\
&= \|\mu_P - \mu_Q\|_{\mathcal{F}}^2
\end{aligned}
$$

$$
\begin{aligned}
\mathbb{E}_P f(x) &= \mathbb{E}_P\left[\langle \varphi_x, f\rangle_{\mathcal{F}}\right] \\
&:= \langle \mu_P, f\rangle_{\mathcal{F}}
\end{aligned}
$$

$$
\mu_P = \mathbb{E}_P \varphi_x = \int k(x,\cdot)dP(x)
$$

$$
\|h\|_{\mathcal{F}} = \sup_{f\in F}\langle f, h\rangle_{\mathcal{F}}
$$

SAFRAN Snecma

ANALYSE DE SENSIBILITÉ GÉNÉRALISÉE

➔ Finalement, Maximum Mean Discrepancy dans le cadre RKHS

$$\begin{aligned}
\mathrm{MMD}^2(P,Q;F) &= \left(\sup_{f\in F}\left[\mathbb{E}_P f(x) - \mathbb{E}_Q f(x)\right]\right)^2 \\
&= \left(\sup_{f\in F}\langle f, \mu_P - \mu_Q\rangle_{\mathcal{F}}\right)^2 \\
&= \|\mu_P - \mu_Q\|^2_{\mathcal{F}}
\end{aligned}$$

$$\begin{aligned}
\mathbb{E}_P f(x) &= \mathbb{E}_P\left[\langle \varphi_x, f\rangle_{\mathcal{F}}\right] \\
&:= \langle \mu_P, f\rangle_{\mathcal{F}}
\end{aligned}$$

$$\mu_P = \mathbb{E}_P \varphi_x = \int k(x,\cdot)dP(x)$$

$$\|h\|_{\mathcal{F}} = \sup_{f\in F}\langle f, h\rangle_{\mathcal{F}}$$

Equivalence entre point de vue fonctionnel et point de vue feature

ANALYSE DE SENSIBILITÉ GÉNÉRALISÉE

➔ **Finalement, Maximum Mean Discrepancy dans le cadre RKHS**

$$
\begin{aligned}
\mathrm{MMD}^2(P,Q;F) &= \left(\sup_{f\in F} \left[\mathbb{E}_P f(x) - \mathbb{E}_Q f(x) \right] \right)^2 \\
&= \left(\sup_{f\in F} \langle f, \mu_P - \mu_Q \rangle_{\mathcal{F}} \right)^2 \\
&= \|\mu_P - \mu_Q\|_{\mathcal{F}}^2
\end{aligned}
$$

$$
\begin{aligned}
\mathbb{E}_P f(x) &= \mathbb{E}_P \left[\langle \varphi_x, f \rangle_{\mathcal{F}} \right] \\
&:= \langle \mu_P, f \rangle_{\mathcal{F}}
\end{aligned}
$$

$$\mu_P = \mathbb{E}_P \varphi_x = \int k(x,\cdot) dP(x)$$

$$\|h\|_{\mathcal{F}} = \sup_{f\in F} \langle f, h \rangle_{\mathcal{F}}$$

Equivalence entre point de vue fonctionnel et point de vue feature

Représentation d'une loi de probabilité dans un RKHS

SAFRAN Snecma

ANALYSE DE SENSIBILITÉ GÉNÉRALISÉE

➔ Finalement, Maximum Mean Discrepancy dans le cadre RKHS

$$\mathrm{MMD}^2(P,Q;F) \quad = \left(\sup_{f \in F} \left[\mathbb{E}_P f(x) - \mathbb{E}_Q f(x) \right] \right)^2$$

➔ Estimateur non biaisé

$$\widehat{\mathrm{MMD}}^2 = \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{j \neq i}^{n} \left[k(x_i, x_j) - k(x_i, x'_j) - k(x'_i, x_j) + k(x'_i, x'_j) \right]$$

$$\{x_i\}_{i=1}^n \sim P, \ \{x'_i\}_{i=1}^n \sim Q$$

SAFRAN Snecma

ANALYSE DE SENSIBILITÉ GÉNÉRALISÉE

➔ Comment comparer des lois de probabilité ?

- Le plus « basique »
  - Comparer leur moyenne

$$d(P_Y, P_{Y|X_i}) = (\mathbb{E}(Y) - \mathbb{E}(Y|X_i))^2$$

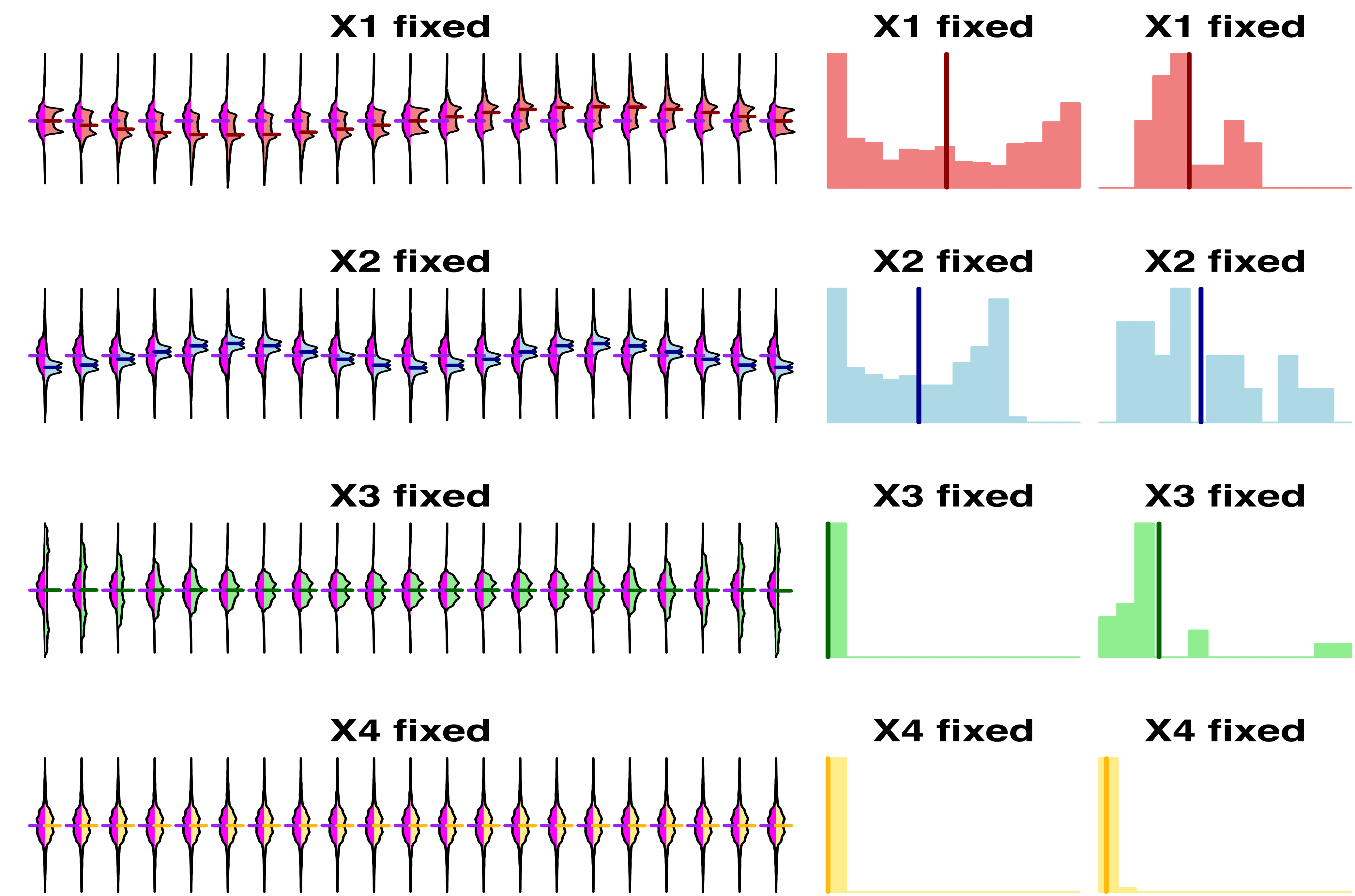
➔ Sobol !

- La famille des f-divergences

$$d_f(P_Y || P_{Y|X_i}) = \int f\left(\frac{p_Y(y)}{p_{Y|X_i}(y)}\right) p_{Y|X_i}(y) dy$$

- MMD version noyau

$$d(P_Y, P_{Y|X_i}) = \mathrm{MMD}^2(Y, Y|X_i)$$


X1 fixed
X1 fixed
X1 fixed
X2 fixed
X2 fixed
X2 fixed
X3 fixed
X3 fixed
X3 fixed
X4 fixed
X4 fixed
X4 fixed

ANALYSE DE SENSIBILITÉ GÉNÉRALISÉE

➔ Note : pourquoi parler d'analyse de sensibilité généralisée ?

- Choix du noyau « libre »
- Pour garantir un indice réellement « distributionnel », il faut un noyau caractéristique
  - Gaussien, Laplace

ANALYSE DE SENSIBILITÉ GÉNÉRALISÉE

➔ Note : pourquoi parler d'analyse de sensibilité généralisée ?

- Choix du noyau « libre »
- Pour garantir un indice réellement « distributionnel », il faut un noyau caractéristique
  - Gaussien, Laplace
- Mais on peut en théorie utiliser le noyau que l'on veut, par exemple le noyau le plus simple qui soit :

$$k(y, y') = \langle y, y' \rangle \overset{1\mathrm{D}}{=} yy'$$

ANALYSE DE SENSIBILITÉ GÉNÉRALISÉE

➔ Note : pourquoi parler d'analyse de sensibilité généralisée ?

- Choix du noyau « libre »
- Pour garantir un indice réellement « distributionnel », il faut un noyau caractéristique
  - Gaussien, Laplace

- Mais on peut en théorie utiliser le noyau que l'on veut, par exemple le noyau le plus simple qui soit :

$$k(y, y') = \langle y, y' \rangle \overset{1\mathrm{D}}{=} yy'$$

$$\Downarrow$$

$$\mathbb{E}\left(\mathrm{MMD}^2(P_Y, P_{Y|X_i})\right) = \mathrm{Var}(\mathbb{E}(Y|X_i))$$

Sobol non normalisé par
la variance de la sortie !

ANALYSE DE SENSIBILITÉ GÉNÉRALISÉE

➔ Bilan à mi-parcours

- Nouveaux indices « distributionnels »
  - Quantification réel de l'impact sur la loi de la sortie
  - Pas de fléau de la dimension pour l'estimation de densité
  - Indices de Sobol comme cas particulier

SAFRAN
Snecma

ANALYSE DE SENSIBILITÉ GÉNÉRALISÉE

➔ Bilan à mi-parcours

- Nouveaux indices « distributionnels »
  - Quantification réel de l'impact sur la loi de la sortie
  - Pas de fléau de la dimension pour l'estimation de densité
  - Indices de Sobol comme cas particulier

➔ Reste un avantage de Sobol

- Décomposition ANOVA
- L'orthogonalité permet de dissocier les effets principaux des interactions
  - Pour certains codes plysiques complexes, les interactions sont très importantes
- Peut-on exhiber une décomposition orthogonale pour des indices distributionnels ?

/03/

DÉCOMPOSITIONS ORTHOGONALES POUR GSA GÉNÉRALISÉE

PLAN

- Contexte et problématique
  - Décomposition de Sobol – ANOVA fonctionnelle
- Analyse de sensibilité généralisée
  - Cadre général à base de densités
  - Vers de nouveaux indices de sensibilité
- **Décompositions orthogonales pour GSA généralisée**
  - Décomposition 1
  - Décomposition 2
  - Le cas des données fonctionnelles
- Conclusions et perspectives

DÉCOMPOSITIONS ORTHOGONALES

➔ **Avant les indices à base de noyaux, distances plus « classiques »**

- Variation totale : aucune décomposition (à ma connaissance)

DÉCOMPOSITIONS ORTHOGONALES

- **Avant les indices à base de noyaux, distances plus « classiques »**
  - Variation totale : aucune décomposition (à ma connaissance)
  - Information mutuelle
    - Une piste dans les travaux sur les mesures d'importance dans les forêts aléatoires

DÉCOMPOSITIONS ORTHOGONALES

➔ Avant les indices à base de noyaux, distances plus « classiques »

- Variation totale : aucune décomposition (à ma connaissance)
- Information mutuelle
  - Une piste dans les travaux sur les mesures d'importance dans les forêts aléatoires

$$\mathrm{MDI}(X_i) = \sum_{k=0}^{d-1} \frac{1}{C_d^k} \frac{1}{d-k} \sum_{B \in P_k(X_{-i})} I(X_i; Y | B)$$

Mean Decrease Impurity
Mesure d'impureté = entropie
Valable pour un ensemble infini d'arbres totalement randomisés et développés avec un échantillon infiniment grand
Entrées et sortie catégorielles

$$\sum_{i=1}^{d} \mathrm{MDI}(X_i) = I(X_1, \ldots, X_d; Y)$$

Louppe et al. 2013

DÉCOMPOSITIONS ORTHOGONALES

➔ **Avant les indices à base de noyaux, distances plus « classiques »**

- Variation totale : aucune décomposition (à ma connaissance)
- Information mutuelle
  - Une piste dans les travaux sur les mesures d'importance dans les forêts aléatoires

- A priori, pas grand chose d'utilisable …

DÉCOMPOSITIONS ORTHOGONALES

➔ Avant les indices à base de noyaux, distances plus « classiques »

- Variation totale : aucune décomposition (à ma connaissance)
- Information mutuelle
  - Une piste dans les travaux sur les mesures d'importance dans les forêts aléatoires

- A priori, pas grand chose d'utilisable …

➔ Est-ce le cas pour les indices MMD ?

- Un espoir car basé sur des moyennes et non sur des densités
- Relativement "proche" de Sobol
  - Rappelez-vous, cas particulier noyau

SAFRAN Snecma

DÉCOMPOSITIONS ORTHOGONALES

➔ Ré-écriture de la décomposition ANOVA

$$\mathrm{Var}(Y) = \sum_{u \subseteq \{1,\ldots,d\}, u \neq \emptyset} g_u$$

$$g_u = \sum_{v \subseteq u} (-1)^{|u|-|v|} \mathrm{Var}\left(\mathbb{E}(Y|X_v)\right)$$

DÉCOMPOSITIONS ORTHOGONALES

➔ Décomposition orthogonale pour indices MMD

$$\mathbb{E}\left(\mathrm{MMD}^2\left(P_{Y|X_{1:d}}, P_Y\right)\right) = \sum_{u \subseteq \{1,\ldots,p\}, u \neq \emptyset} g_u$$

$$g_u = \sum_{v \subseteq u} (-1)^{|u|-|v|} \mathbb{E}\left(\mathrm{MMD}^2\left(P_{Y|X_v}, P_Y\right)\right)$$

- Preuve : combinatoire

$$P_{Y|X_{1:d}} - P_Y = \sum_{u \subseteq \{1,\ldots,d\}, u \neq \emptyset} \sum_{v \subseteq u} (-1)^{|u|-|v|} P_{Y|X_v}$$

$$\mathbb{E}\left(\left\langle \sum_{v \subseteq u} (-1)^{|u|-|v|} P_{Y|X_v}, \sum_{v \subseteq u'} (-1)^{|u|-|v|} P_{Y|X_v} \right\rangle_k\right) = 0$$

DÉCOMPOSITIONS ORTHOGONALES

➔ Décomposition orthogonale pour indices MMD

$$\mathbb{E}\left(\mathrm{MMD}^2\left(P_{Y|X_{1:d}}, P_Y\right)\right) = \sum_{u\subseteq\{1,\dots,p\}, u\neq\emptyset} g_u$$

$$g_u = \sum_{v\subseteq u} (-1)^{|u|-|v|}\mathbb{E}\left(\mathrm{MMD}^2\left(P_{Y|X_v}, P_Y\right)\right)$$

$$\boxed{S_u^{\mathrm{MMD}} = \frac{\sum_{v\subseteq u}(-1)^{|u|-|v|}\mathbb{E}\left(\mathrm{MMD}^2\left(P_{Y|X_v}, P_Y\right)\right)}{\mathbb{E}\left(\mathrm{MMD}^2\left(P_{Y|X_{1:d}}, P_Y\right)\right)}}$$

INTERLUDE : ESTIMATION

➔ **Les indices de Sobol et MMD évitent l'estimation de densité**

- Déjà vu, avantage vs autres indices distributionnels

INTERLUDE : ESTIMATION

➔ **Les indices de Sobol et MMD évitent l'estimation de densité**

- Déjà vu, avantage vs autres indices distributionnels

➔ **MAIS ils font malgré tout intervenir les lois conditionnelles**

$$d(P_Y, P_{Y|X_i}) = (\mathbb{E}(Y) - \mathbb{E}(Y|X_i))^2$$

$$d(P_Y, P_{Y|X_i}) = \mathrm{MMD}^2(Y, Y|X_i)$$

- Coûteux, d'où l'utilisation en pratique de modèles approchés (émulateur, proxy, …)

INTERLUDE : ESTIMATION

➔ **Les indices de Sobol et MMD évitent l'estimation de densité**

- Déjà vu, avantage vs autres indices distributionnels

➔ **MAIS ils font malgré tout intervenir les lois conditionnelles**

$$d(P_Y, P_{Y|X_i}) = (\mathbb{E}(Y) - \mathbb{E}(Y|X_i))^2$$

$$d(P_Y, P_{Y|X_i}) = \mathrm{MMD}^2(Y, Y|X_i)$$

- Coûteux, d'où l'utilisation en pratique de modèles approchés (émulateur, proxy, …)

➔ **Peut-on s'inspirer du cas de l'information mutuelle ?**

- Estimation de densité certes, mais pas de lois conditionnelles (uniquement lois jointes)

SAFRAN Snecma

INTERLUDE : MESURE DE DÉPENDANCE ENTRE VARIABLES ALÉATOIRES

➔ Comment comparer des lois de probabilité ?

- Le plus « basique »
  - Comparer leur moyenne

$$d(P_Y, P_{Y|X_i}) = (\mathbb{E}(Y) - \mathbb{E}(Y|X_i))^2$$

➔ Sobol !

- La famille des f-divergences

$$d_f(P_Y || P_{Y|X_i}) = \int f\left(\frac{p_Y(y)}{p_{Y|X_i}(y)}\right) p_{Y|X_i}(y)dy$$

  - Inclut comme cas particuliers d'indices de sensibilité la distance TV et l'information mutuelle

$$S_i^{TV} = \int |p_Y(y) - p_{Y|X_i=x}(y)| p_{X_i}(x)dxdy$$

Borgonovo 2007

$$\boxed{S_i^{KL} = \int p_{Y|X_i=x}(y) \ln\left(\frac{p_{Y|X_i=x}(y)}{p_Y(y)}\right) p_{X_i}(x)dxdy}$$

INTERLUDE : MESURE DE DÉPENDANCE ENTRE VARIABLES ALÉATOIRES

➔ **L'information mutuelle est une mesure de la dépendance entre deux vecteurs aléatoires**

$$
\begin{aligned}
S_i^{KL} &= \int p_{Y|X_i=x}(y) \ln\left(\frac{p_{Y|X_i=x}(y)}{p_Y(y)}\right) p_{X_i}(x)dxdy \\
&= \int p_{Y,X_i}(y,x) \ln\left(\frac{p_{Y,X_i}(y,x)}{p_Y(y)p_{X_i}(x)}\right) dxdy = I(X_i;Y)
\end{aligned}
$$

INTERLUDE : MESURE DE DÉPENDANCE ENTRE VARIABLES ALÉATOIRES

➔ **L'information mutuelle est une mesure de la dépendance entre deux vecteurs aléatoires**

$$
\begin{aligned}
S_i^{KL} &= \int p_{Y|X_i=x}(y) \ln\left(\frac{p_{Y|X_i=x}(y)}{p_Y(y)}\right) p_{X_i}(x) dx dy \\
&= \int p_{Y,X_i}(y,x) \ln\left(\frac{p_{Y,X_i}(y,x)}{p_Y(y) p_{X_i}(x)}\right) dx dy = I(X_i;Y)
\end{aligned}
$$

➔ **On voit donc qu'un indice de sensibilité peut dans certains cas s'interpréter comme une mesure de dépendance entre entrées et sortie**

- Question : l'information mutuelle souffrant du fléau de la dimension, peut-on envisager d'autres mesures de dépendance ?
- Subsidiaire : avec une décomposition orthogonale ?

INTERLUDE : MESURE DE DÉPENDANCE ENTRE VARIABLES ALÉATOIRES

➔ **En toute généralité, une mesure de dépendance compare la loi jointe et le produit des marginales**

- Si proche, les variables sont dépendantes
- Comment comparer loi jointe et produit des marginales ?
  - Sans estimation de densité, une idée ?

INTERLUDE : MESURE DE DÉPENDANCE ENTRE VARIABLES ALÉATOIRES

- **En toute généralité, une mesure de dépendance compare la loi jointe et le produit des marginales**
  - Si proche, les variables sont dépendantes
  - Comment comparer loi jointe et produit des marginales ?
    - Sans estimation de densité, une idée ?

$$\mathrm{MMD}^2\left(P_{Y,X}, P_Y P_X\right)$$

SAFRAN Snecma

INTERLUDE : MESURE DE DÉPENDANCE ENTRE VARIABLES ALÉATOIRES

- **En toute généralité, une mesure de dépendance compare la loi jointe et le produit des marginales**
  - Si proche, les variables sont dépendantes
  - Comment comparer loi jointe et produit des marginales ?
    - Sans estimation de densité, une idée ?

$$\mathrm{MMD}^2\left(P_{Y,X}, P_Y P_X\right) = \left(\sup_{f\in F}\left[\mathbb{E}_{P_{XY}} f(x,y) - \mathbb{E}_{P_X P_Y} f(x,y)\right]\right)^2$$

INTERLUDE : MESURE DE DÉPENDANCE ENTRE VARIABLES ALÉATOIRES

- **En toute généralité, une mesure de dépendance compare la loi jointe et le produit des marginales**
  - Si proche, les variables sont dépendantes
  - Comment comparer loi jointe et produit des marginales ?
    - Sans estimation de densité, une idée ?

$$
\begin{aligned}
\mathrm{MMD}^2\left(P_{Y,X}, P_Y P_X\right) &= \left(\sup_{f\in F}\left[\mathbb{E}_{P_{XY}} f(x,y) - \mathbb{E}_{P_X P_Y} f(x,y)\right]\right)^2 \\
&= \left\|\mu_{P_{XY}} - \mu_{P_X P_Y}\right\|^2_{\mathcal{F}\times\mathcal{G}}
\end{aligned}
$$

INTERLUDE : MESURE DE DÉPENDANCE ENTRE VARIABLES ALÉATOIRES

➔ **En toute généralité, une mesure de dépendance compare la loi jointe et le produit des marginales**

- Si proche, les variables sont dépendantes
- Comment comparer loi jointe et produit des marginales ?
  - Sans estimation de densité, une idée ?

$$
\begin{aligned}
\mathrm{MMD}^2\left(P_{Y,X}, P_Y P_X\right) &= \left(\sup_{f \in F}\left[\mathbb{E}_{P_{XY}} f(x,y) - \mathbb{E}_{P_X P_Y} f(x,y)\right]\right)^2 \\
&= \left\|\mu_{P_{XY}} - \mu_{P_X P_Y}\right\|^2_{\mathcal{F}\times\mathcal{G}} \\
&= \mathrm{HSIC}(X, Y)
\end{aligned}
$$

Hilbert-Schmidt Independence Criterion
Gretton 2005

INTERLUDE : MESURE DE DÉPENDANCE ENTRE VARIABLES ALÉATOIRES

➔ **En toute généralité, une mesure de dépendance compare la loi jointe et le produit des marginales**

- Si proche, les variables sont dépendantes
- Comment comparer loi jointe et produit des marginales ?
  - Sans estimation de densité, une idée ?

$$\begin{aligned} \mathrm{MMD}^2\left(P_{Y,X}, P_Y P_X\right) &= \left(\sup_{f \in F} \left[\mathbb{E}_{P_{XY}} f(x,y) - \mathbb{E}_{P_X P_Y} f(x,y)\right]\right)^2 \\ &= \|\mu_{P_{XY}} - \mu_{P_X P_Y}\|^2_{\mathcal{F}\times\mathcal{G}} \\ &= \mathrm{HSIC}(X,Y) \end{aligned}$$

Hilbert-Schmidt Independence Criterion
Gretton 2005

➔ **La mesure HSIC dépend d'un noyau défini dans l'espace joint**

- En pratique, produit de noyaux définis sur chaque vecteur
- Estimation ne faisant pas intervenir les lois conditionnelles !

INTERLUDE : MESURE DE DÉPENDANCE ENTRE VARIABLES ALÉATOIRES

➔ Estimation de la mesure HSIC

$$\widehat{\mathrm{HSIC}}(X,Y) = \frac{1}{n^2}\mathrm{trace}\,(KHLH)$$

$$[K]_{ij} = k_{\mathcal{X}}(x_i, x_j) \qquad [L]_{ij} = k_{\mathcal{Y}}(y_i, y_j) \qquad [H]_{ij} = \delta_{ij} - \frac{1}{n}$$

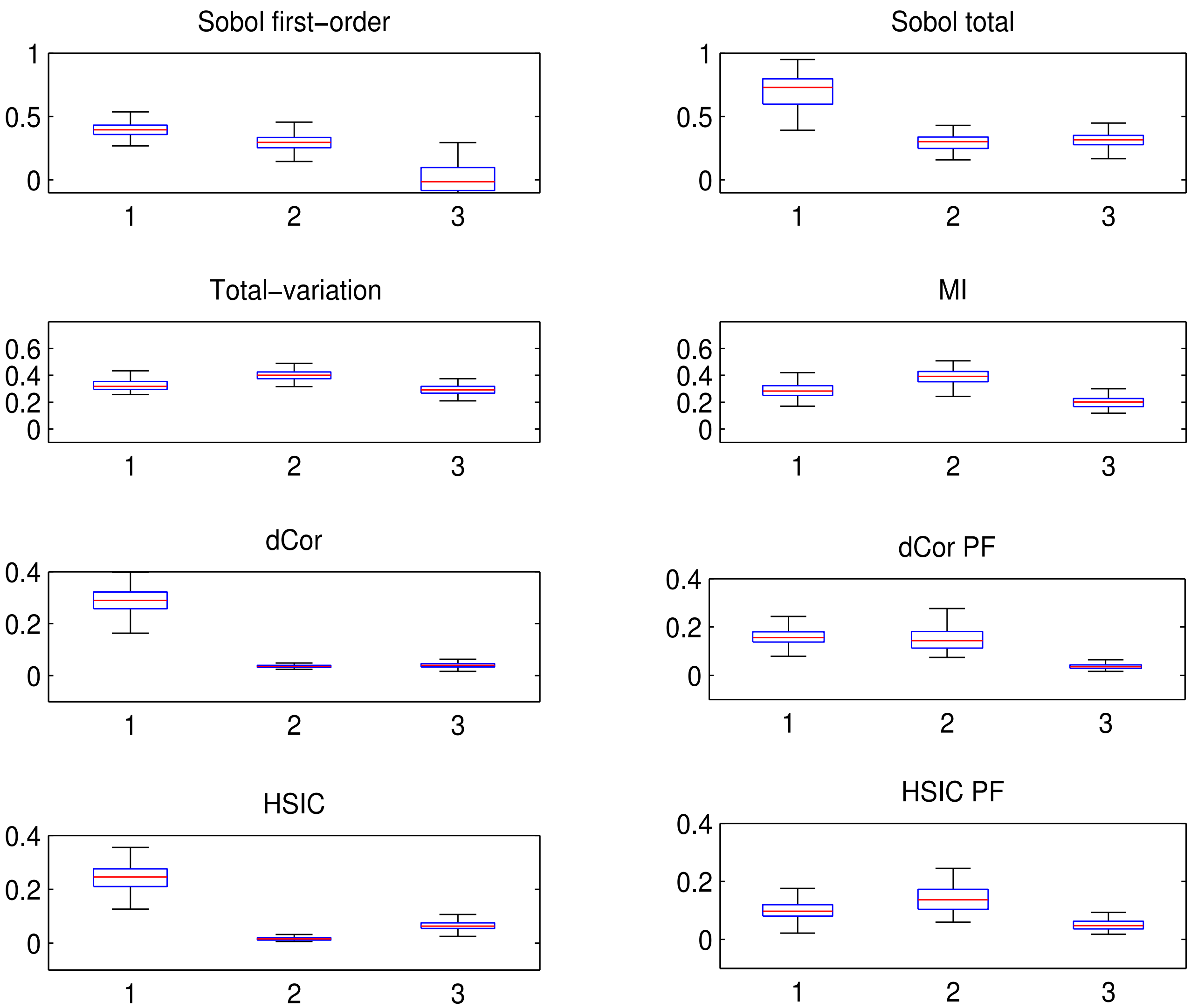
$$\{(x_i, y_i)\}_{i=1}^{n} \sim P_{XY}$$

INTERLUDE : MESURE DE DÉPENDANCE ENTRE VARIABLES ALÉATOIRES

➔ **Une version normalisée de cette mesure de dépendance a été étudiée comme indice de sensibilité et comparée aux indices de Sobol (D. 2014)**

- Beaucoup moins d'observations nécessaires que pour les indices de Sobol
- Très bonnes propriétés de screening

INTERLUDE : MESURE DE DÉPENDANCE ENTRE VARIABLES ALÉATOIRES



D. 2014

INTERLUDE : MESURE DE DÉPENDANCE ENTRE VARIABLES ALÉATOIRES

➔ **Une version normalisée de cette mesure de dépendance a été étudiée comme indice de sensibilité et comparée aux indices de Sobol (D. 2014)**

- Beaucoup moins d'observations nécessaires que pour les indices de Sobol
- Très bonnes propriétés de screening

➔ **Mais ce n'était qu'un travail préliminaire**

- La normalisation proposée n'était en fait pas la bonne !
- Pas de décomposition ANOVA …

SAFRAN
Snecma

INTERLUDE : MESURE DE DÉPENDANCE ENTRE VARIABLES ALÉATOIRES

- **Une version normalisée de cette mesure de dépendance a été étudiée comme indice de sensibilité et comparée aux indices de Sobol (D. 2014)**
  - Beaucoup moins d'observations nécessaires que pour les indices de Sobol
  - Très bonnes propriétés de screening

- **Mais ce n'était qu'un travail préliminaire**
  - La normalisation proposée n'était en fait pas la bonne !
  - Pas de décomposition ANOVA …
  - … jusqu'à aujourd'hui
    - Décomposition orthogonale, mais pas dans le cas général (hypothèse sur noyau entrées)
    - Fournit donc la bonne normalisation comme pour Sobol et MMD

SAFRAN Snecma

DÉCOMPOSITIONS ORTHOGONALES

➔ Ré-écriture de la décomposition ANOVA

$$\mathrm{Var}(Y) = \sum_{u \subseteq \{1,\dots,d\}, u \neq \emptyset} g_u$$

$$g_u = \sum_{v \subseteq u} (-1)^{|u|-|v|} \mathrm{Var}\left(\mathbb{E}(Y|X_v)\right)$$

DÉCOMPOSITIONS ORTHOGONALES

➔ Décomposition orthogonale pour indices HSIC

$$\mathrm{HSIC}\left(Y, X_{1:d}\right) = \sum_{u \subseteq \{1,\ldots,p\}, u \neq \emptyset} g_u$$

$$g_u = \sum_{v \subseteq u} (-1)^{|u|-|v|} \mathrm{HSIC}\left(Y, X_v\right)$$

- Sous l'hypothèse que le noyau sur chaque entrée vérifie

$$\int_{\mathcal{X}} k_{\mathcal{X}}(x, x') dP_X(x) = 1$$

$$\begin{aligned}
k_{\mathcal{X}} &= 1 + k^0_{\mathcal{X}} \\
k^0_{\mathcal{X}}(x, x') &= k(x, x') - \frac{\int_{\mathcal{X}} k(x, x') dP_X(x') \int_{\mathcal{X}} k(x, x') dP_X(x)}{\iint_{\mathcal{X}\times\mathcal{X}} k(x, x') dP_X(x) dP_X(x')} \\
k^0_{\mathcal{X}}(x, x') &= k(x, x') - \int_{\mathcal{X}} k(x, x') dP_X(x') - \int_{\mathcal{X}} k(x, x') dP_X(x) \\
&+ \iint_{\mathcal{X}\times\mathcal{X}} k(x, x') dP_X(x) dP_X(x')
\end{aligned}$$

DÉCOMPOSITIONS ORTHOGONALES

➔ Décomposition orthogonale pour indices HSIC

$$\mathrm{HSIC}\left(Y, X_{1:d}\right) = \sum_{u \subseteq \{1,\ldots,p\}, u \neq \emptyset} g_u$$

$$g_u = \sum_{v \subseteq u} (-1)^{|u|-|v|} \mathrm{HSIC}\left(Y, X_v\right)$$

$$\boxed{S_u^{\mathrm{HSIC}} = \frac{\sum_{v \subseteq u} (-1)^{|u|-|v|} \mathrm{HSIC}\left(Y, X_v\right)}{\mathrm{HSIC}\left(Y, X_{1:d}\right)}}$$

DÉCOMPOSITIONS ORTHOGONALES

➔ **Note le retour : pourquoi parler d'analyse de sensibilité généralisée ?**

- Choix du noyau « libre »
- Pour garantir un indice réellement « distributionnel », il faut un noyau caractéristique
  - Gaussien, Laplace
- Pour garantir la décomposition orthogonale, hypothèse sur noyau entrées

DÉCOMPOSITIONS ORTHOGONALES

➔ Note le retour : pourquoi parler d'analyse de sensibilité généralisée ?

- Choix du noyau « libre »
- Pour garantir un indice réellement « distributionnel », il faut un noyau caractéristique
  - Gaussien, Laplace
- Pour garantir la décomposition orthogonale, hypothèse sur noyau entrées

- Si entrées uniformes et noyau tend vers le « noyau » Dirac

$$k_{\mathcal{X}}(x, x') \to \delta(x, x')$$

$$\text{ex: } k_{\mathcal{X}}(x, x') = \frac{1}{\sqrt{2\pi a^2}} \exp\left(\frac{1}{2a^2}(x - x')^2\right), \quad a \to 0$$

DÉCOMPOSITIONS ORTHOGONALES

➔ **Note le retour : pourquoi parler d'analyse de sensibilité généralisée ?**

- Choix du noyau « libre »
- Pour garantir un indice réellement « distributionnel », il faut un noyau caractéristique
  - Gaussien, Laplace
- Pour garantir la décomposition orthogonale, hypothèse sur noyau entrées

- Si entrées uniformes et noyau tend vers le « noyau » Dirac

$$k_{\mathcal{X}}(x, x') \rightarrow \delta(x, x')$$

ex: $k_{\mathcal{X}}(x, x') = \frac{1}{\sqrt{2\pi a^2}} \exp\left(\frac{1}{2a^2}(x - x')^2\right), \quad a \rightarrow 0$

$$\boxed{S_u^{\mathrm{HSIC}} \longrightarrow S_u^{\mathrm{MMD}}}$$

DÉCOMPOSITIONS ORTHOGONALES

➔ **Note le retour : pourquoi parler d'analyse de sensibilité généralisée ?**

- Choix du noyau « libre »
- Pour garantir un indice réellement « distributionnel », il faut un noyau caractéristique
  - Gaussien, Laplace
- Pour garantir la décomposition orthogonale, hypothèse sur noyau entrées

- Si entrées uniformes et noyau tend vers le « noyau » Dirac

$$k_{\mathcal{X}}(x, x') \to \delta(x, x')$$

$$\text{ex: } k_{\mathcal{X}}(x, x') = \frac{1}{\sqrt{2\pi a^2}} \exp\left(\frac{1}{2a^2}(x - x')^2\right), \quad a \to 0$$

$$\boxed{S_u^{\mathrm{HSIC}} \longrightarrow S_u^{\mathrm{MMD}}}$$

$$k_{\mathcal{Y}}(y, y') = yy' \qquad \boxed{S_u^{\mathrm{MMD}} = S_u^{\mathrm{Sobol}}}$$

DÉCOMPOSITIONS ORTHOGONALES
VOUS SOUVENEZ-VOUS DU MENU ?

➔ **Résumé de la démarche pratique :**

1. Screening pour limiter le nombre de paramètres
2. Analyse quantitative pour hiérarchiser finement ceux qui restent

➔ **Mais pour des problèmes industriels sérieux**

- Coût calculatoire important du modèle numérique
  - Même les estimateurs state-of-the-art sont impossibles à utilis
- La dimension du vecteur des entrées peut être très grande
  - Techniques de screening pas forcément robustes
- La variance de la sortie n'est pas forcément la quantité d'int
  - Autres indices de sensibilité ? (e.g. pour probabilité dépassem
- Les entrées et les sorties ne sont pas forcément des scalai
  - Courbes d'évolution, maillages 3D, …

Un cadre général ?

Emulateurs

Feature selection

Nvx indices

Distances, décompositions

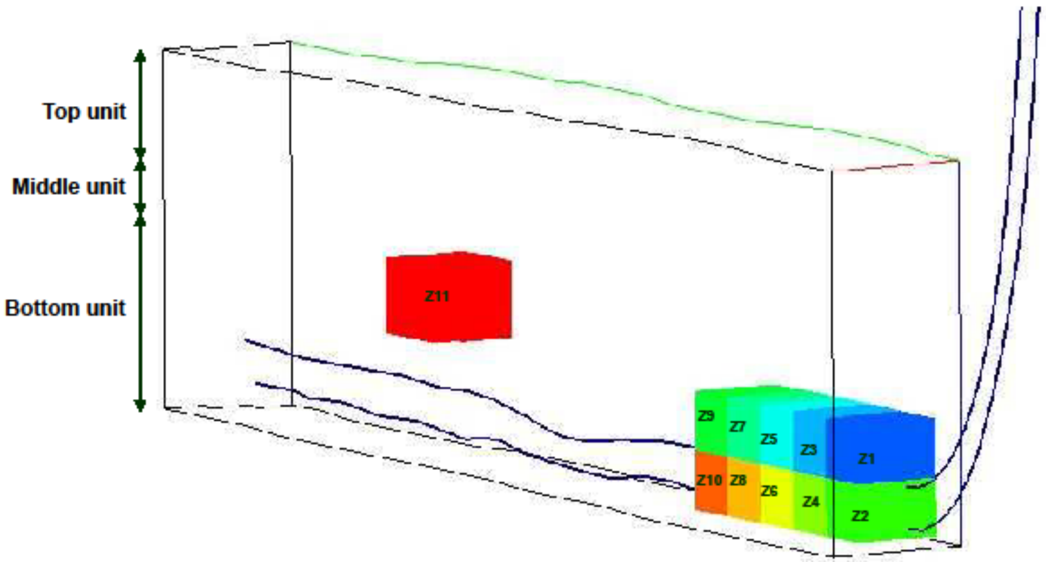
SAFRAN Snecma

DÉCOMPOSITIONS ORTHOGONALES

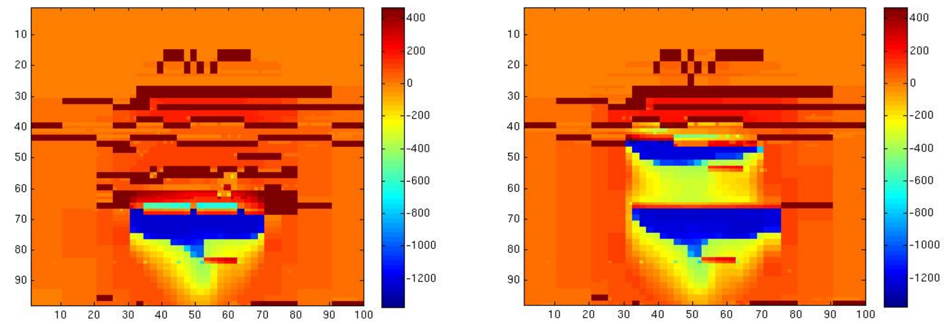
➔ Prise en compte d'entrées/sorties fonctionnelles

- Dans la « vraie » vie, cela ressemble à ça

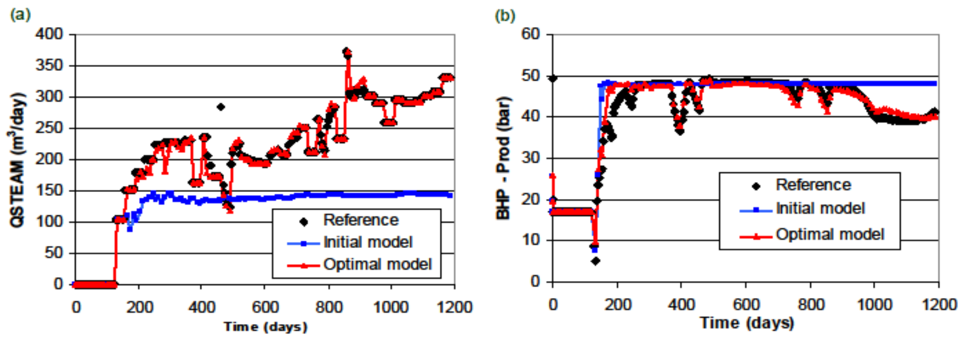
Réservoir de pétrole produit par injection de vapeur



Entrées : proportions d'argile dans certaines zones



Sortie : cube 3D de propagation de la chambre de vapeur dans le réservoir



Sortie : évolution temporelle du débit de vapeur et de la pression au puits producteur

SAFRAN Snecma

DÉCOMPOSITIONS ORTHOGONALES

➔ Prise en compte d'entrées/sorties fonctionnelles

- Dans la « vraie » vie, cela ressemble à ça

- Pour indices de Sobol, généralisation pour un vecteur de sorties (Gamboa et al. 2013)
  - Combinaison linéaire des indices de chaque composante
  - Parmi les mesures basées sur la variance, seule façon d'avoir des propriétés d'invariance

- L'approche par noyaux permet un point de vue complètement nouveau
  - En recyclant les (très nombreux) travaux déjà effectués pour construire des noyaux adaptés dans plusieurs applications !
    - Ex: noyaux pour l'analyse d'images
    - Ex: noyaux pour le texte (sait-on jamais !)
    - Ex: noyaux pour séries temporelles
    - Etc …

SAFRAN Snecma

DÉCOMPOSITIONS ORTHOGONALES

➔ Prise en compte d'entrées/sorties fonctionnelles

- Dans la « vraie » vie, cela ressemble à ça

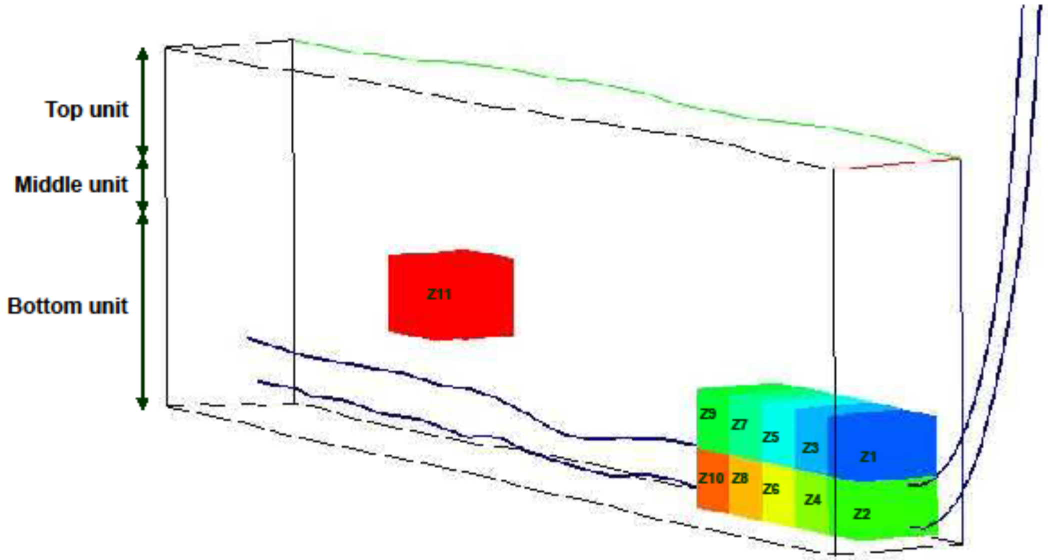
- Pour indices de Sobol, généralisation pour un vecteur de sorties (Gamboa et al. 2013)
  - Combinaison linéaire des indices de chaque composante
  - Parmi les mesures basées sur la variance, seule façon d'avoir des propriétés d'invariance

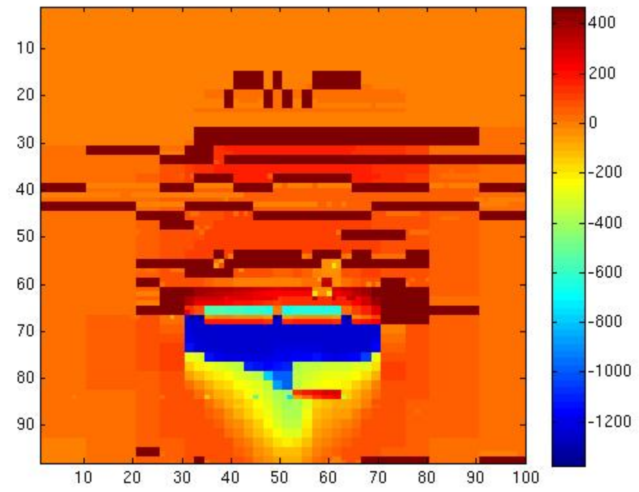
- L'approche par noyaux permet un point de vue complètement nouveau
  - En recyclant les (très nombreux) travaux déjà effectués pour construire des noyaux adaptés dans plusieurs applications !
    - Ex: noyaux pour l'analyse d'images
    - Ex: noyaux pour le texte (sait-on jamais !)
    - Ex: noyaux pour séries temporelles
    - Etc …

- Exemple sur le cas du réservoir de pétrole avec un cube 3D de propriétés physiques en sortie

DÉCOMPOSITIONS ORTHOGONALES

➔ **Modèle de simulation d'écoulement de pétrole dans un réservoir de pétrole produit par injection de vapeur d'eau (sables bitumineux Canada)**



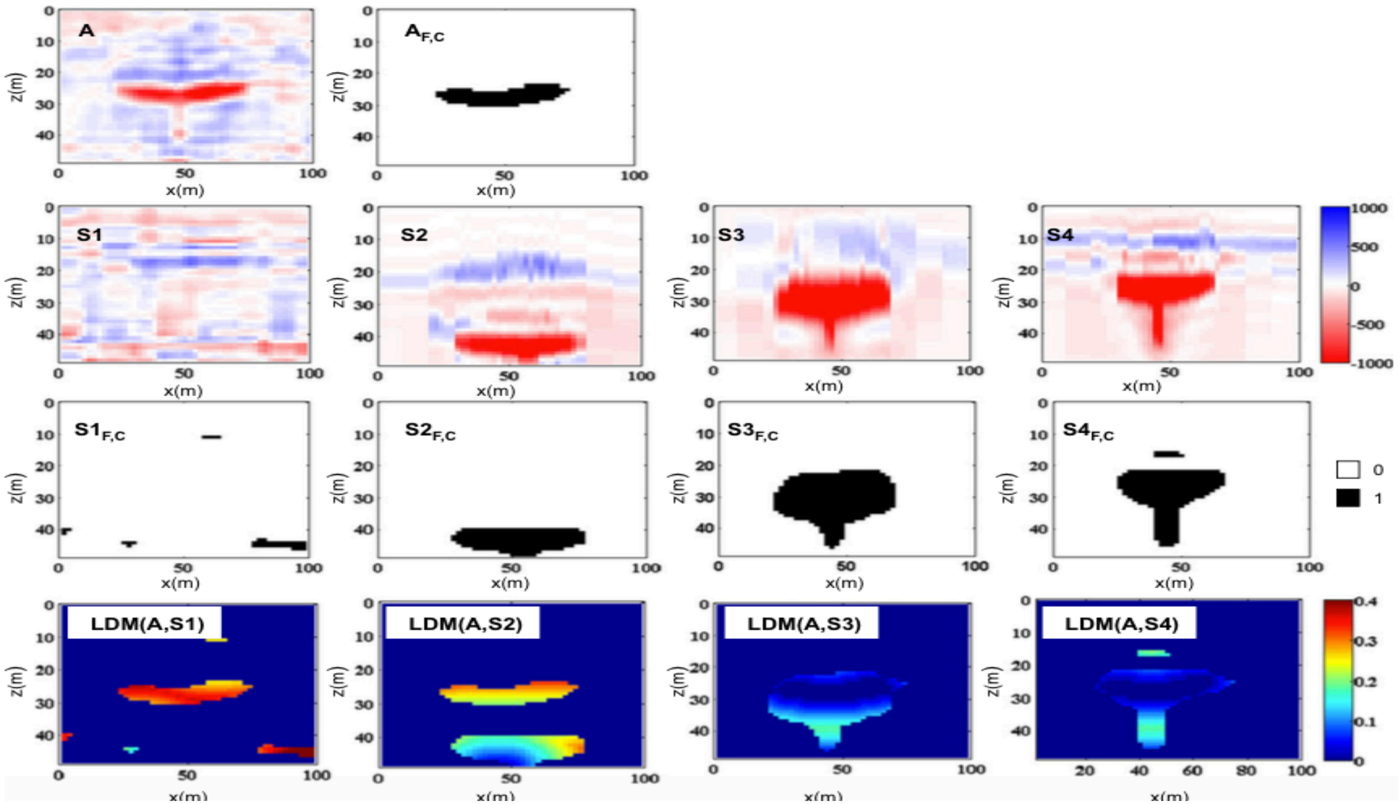


Entrées : 11 proportions d'argile dans certaines zones du réservoir

Sortie : cube 3D d'impédances (rouge = argile, bleu = vapeur)
Permet de localiser le panache de vapeur et vérifier que toutes les régions du réservoir sont chauffées

DÉCOMPOSITIONS ORTHOGONALES

- **Modèle de simulation d’écoulement de pétrole dans un réservoir de pétrole produit par injection de vapeur d’eau (sables bitumineux Canada)**
- **Noyau basé sur une mesure de similarité entre cubes 3D adaptée au problème**



Tillier et al. 2013

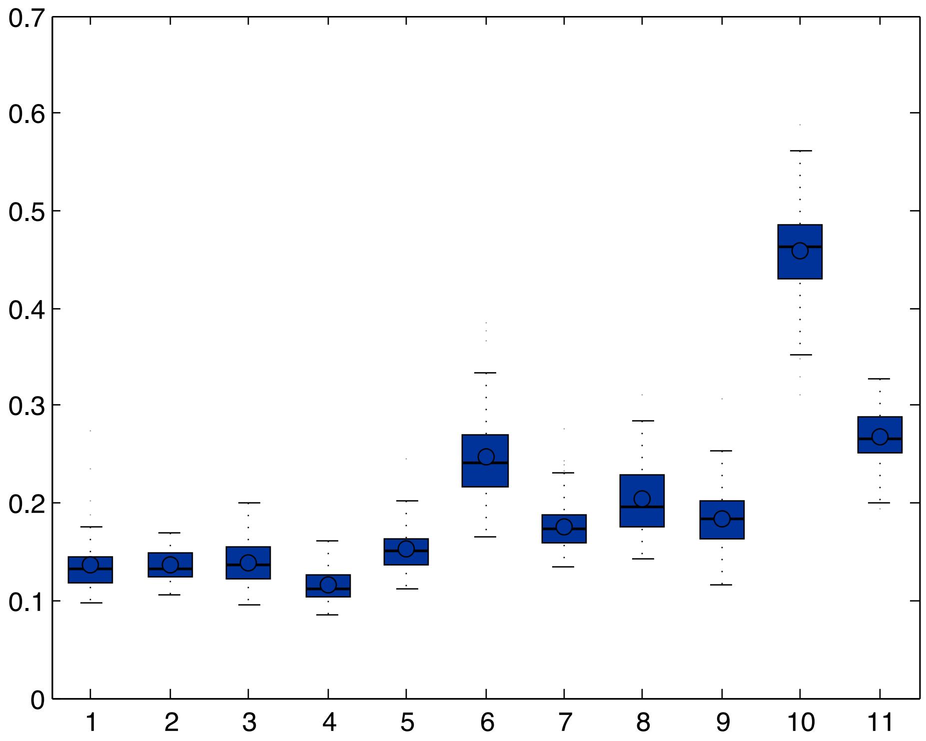
DÉCOMPOSITIONS ORTHOGONALES

- **Modèle de simulation d'écoulement de pétrole dans un réservoir de pétrole produit par injection de vapeur d'eau (sables bitumineux Canada)**
- **Noyau basé sur une mesure de similarité entre cubes 3D adaptée au problème**
  - Et comparaison avec un autre noyau ACP
- **Analyse de sensibilité à l'aide de la mesure HSIC**
  - Echantillon de taille 100
  - Bootstrap pour évaluer la variabilité de l'estimation

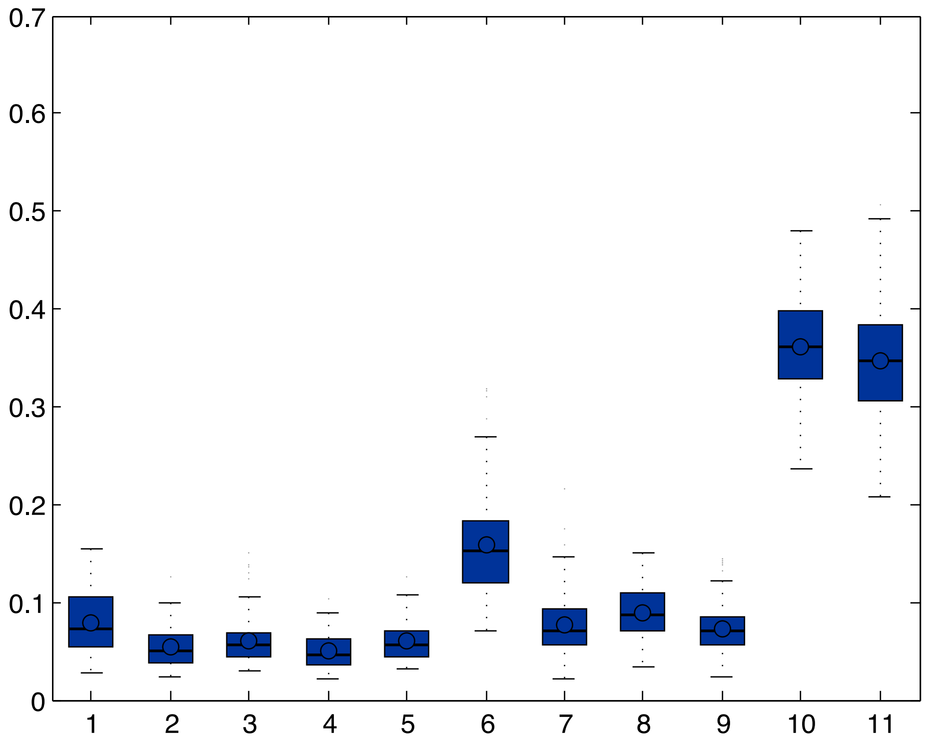
SAFRAN Snecma

DÉCOMPOSITIONS ORTHOGONALES

Noyau ACP



Noyau LDM



SAFRAN Snecma

/04/

CONCLUSIONS & PERSPECTIVES

CONCLUSION

➔ Nouvelle perspective GSA

- Consistante, englobe indices de Sobol mais plus générale
- Très proche du machine learning finalement
  - Recyclage noyaux
- La complexité des entrées ou des sorties est transparente
  - Une fois les noyaux choisis
  - Richesse au-delà du simple noyau produit scalaire (Sobol)
- Logiciels
  - f-divergences + HSIC (mauvaise normalisation) disponibles dans le package R **sensitivity**

➔ Les sujets d'intérêt à l'heure actuelle :

- Choix du noyau
- Travaux intensifs de comparaison sur plusieurs cas test vs Sobol
  - Notamment, analyse des différences de détection

CONCLUSION

➔ **Communauté française : GDR Mascot-Num**

- Ateliers, conférence annuelle
- News, bibliographie, formations

Groupement de Recherche
MASCOT-NUM
Méthodes d'Analyse Stochastique pour les COdes et Traitements NUMériques

Merci de votre attention !

Questions ?

SAFRAN
Snecma