

# LIENS ENTRE DISCRÉPANCE ET ESTIMATION NON-PARAMÉTRIQUE, MÉTHODOLOGIE DE SÉLECTION DE POINTS SELON LES DONNÉES DISPONIBLES

Vincent FEUILLARD<sup>1 2</sup>

## RÉSUMÉ

L'objectif de cet article est d'évaluer la qualité d'une base de données représentant la manière dont celle-ci occupe au mieux son domaine de variation. Le travail réalisé ici propose des outils mathématiques et algorithmiques permettant de réaliser une telle opération. Des techniques d'extraction et d'importation de nouvelles observations sont étudiées. Leurs applications seront illustrées dans le cadre de l'évaluation de paramètres fonctionnels dans un contexte d'estimation par fonctions orthogonales.

*Mots-clés* : « Space filling designs », discrédance, estimation fonctionnelle.

## ABSTRACT

This article aims at evaluating the quality of a data base by checking that the data, with respect to our objectives, « best fill » the space. This work provides some algorithmic and mathematical tools to achieve such a purpose. Extraction and importation techniques to improve the global quality of the data are proposed. An illustration of its application is exposed in the context of functional estimation with orthogonal functions.

*Keywords* : Space filling design, discrepancy, functional estimation.

## 1. Introduction

Dans cet article nous nous intéressons aux liens qui existent entre une méthode d'estimation d'un paramètre fonctionnel et un critère d'uniformité d'un ensemble de points appelé discrédance. Nous complétons et développons ici l'article de Feuillard *et al.* (2006). Une méthodologie ayant pour objectif la

---

1. LSTA, Université Pierre et Marie Curie, Paris VI, 175 rue du Chevaleret, 75013 Paris, CEA Cadarache, DEN/DER/SESI/LCFR, 13108 Saint-Paul-les-Durance Cedex.

2. Adresse actuelle : EADS Innovation Work, Applied Mathematic Group, 12 rue Pasteur, 92150 Suresnes Cedex.

construction ou l'amélioration d'une base de données en vue d'une meilleure estimation (au sens des critères d'IMSE, *Integrated Mean Square Error* et de MSE, *Mean Square Error*, définis plus loin) d'un paramètre fonctionnel en sera déduite.

Le modèle considéré est le suivant. On observe les  $y_i$ ,  $i = 1 \dots, n$ , où :

$$y_i = f(x_i) + \varepsilon_i, \quad \text{où,} \quad (1)$$

- $x_i \in \mathcal{X} = [0, 1]^d$ ,
- $f(\cdot) \in L^2(\mathcal{X})$  est le paramètre fonctionnel à estimer,
- $\varepsilon_i$ ,  $i = 1, \dots, n$ , est une suite de variables aléatoires, mutuellement indépendantes de moyenne nulle de variance  $\sigma^2$ , et indépendantes des  $x_i$ ,  $i = 1, \dots, n$ .

La méthode d'estimation fonctionnelle utilisée, introduite par Cencov (1962), consiste tout d'abord à projeter  $f(\cdot)$  sur un sous-espace de dimension finie de l'espace de Hilbert des fonctions de carré intégrable  $(L^2(\mathcal{X}), \|\cdot\|_{L^2})$ . Une estimation de la projection de  $f(\cdot)$  sur ce sous-espace est alors obtenue par une méthode de moments empiriques. Cette méthode, aujourd'hui classique, et dite *des fonctions orthogonales*, a fait l'objet de nombreuses études, notamment dans le cas de l'estimation d'une densité de probabilité. Nous renvoyons, par exemple, à Bosq (1969), Sansone (1977), Prakasa Rao (1983), et aux ouvrages de Devroye et Györfi (1985), Bosq et Lecoutre (1987), Härdle (1989), ainsi que leurs références bibliographiques.

La fonction  $f(\cdot)$  étant supposée de carré intégrable sur  $\mathcal{X}$ , celle-ci peut s'écrire à partir d'un système de fonctions orthonormales  $\{v_1(\cdot), \dots, v_k(\cdot), \dots\} \subset L^2(\mathcal{X})$ , soit

$$f(x) = \sum_{k \geq 1} a_k v_k(x) \quad \text{où} \quad a_k := \int_{\mathcal{X}} f(x) v_k(x) dx. \quad (2)$$

L'estimation de  $f(\cdot)$ , par projection sur le sous-espace défini par les  $N$  premières fonctions du système orthonormal  $\{v_1(\cdot), \dots, v_k(\cdot), \dots\}$ , s'écrit alors sous la forme :

$$\hat{f}_n(x) := \sum_{k=1}^N \hat{a}_k v_k(x) \quad \text{où} \quad \hat{a}_k := \frac{1}{n} \sum_{i=1}^n y(x_i) v_k(x_i). \quad (3)$$

Les coefficients  $\hat{a}_k$  sont des estimateurs empiriques sans biais des  $a_k$ . Pour le choix de la dimension du sous-espace  $N$  en fonction de la taille  $n$  de l'échantillon, nous renvoyons, par exemple, à Bosq et Bluez (1978), et Aubin (2005), ainsi qu'à leurs références bibliographiques. Il est classique de supposer que, lorsque  $n \rightarrow \infty$ , on a  $N/n \rightarrow 0$  et  $N \rightarrow \infty$ .

Pour ce type d'estimation, Hickernell (1999) et Rafajłowicz et Schwabe (2006) ont montré qu'il existe un lien entre un critère d'uniformité d'un ensemble de points  $\mathbf{x}(n) = \{x_1, \dots, x_n\}$  dans l'espace  $\mathcal{X} = [0, 1]^d$ , appelé discrédance, et des critères de qualité d'estimation de  $f(\cdot)$ . Nous retiendrons ici les critères

de la MSE (Mean Square Error), et de l'IMSE, *Integrated Mean Square Error* (voir l'équation (11) du paragraphe 2) et les équations (13) du §3 pour des définitions précises de ces quantités).

Dans la suite de cet article, nous rappellerons tout d'abord l'inégalité de Koksma-Hlwaka généralisée, faisant intervenir la discrétance généralisée. Cette inégalité sera ensuite utilisée pour fournir des majorations des critères d'IMSE et de MSE. A partir de ces résultats, une méthodologie d'analyse, de sélection et de spécification de points expérimentaux sera proposée. Enfin, une application à un exemple illustrera cette méthodologie.

## 2. Discrétance et inégalité de Koksma-Hlwaka généralisée

### 2.1. Notations et hypothèse

Avant de définir l'inégalité de Koksma-Hlwaka généralisée, commençons par préciser quelques notations.

- i) On pose  $\mathcal{X} = [0, 1]^d$ ;
- ii)  $\mathbf{u} = \{u_1, \dots, u_\ell\}$  désigne un sous-ensemble non vide d'indices de  $\{1, \dots, d\}$ ;
- iii)  $|\mathbf{u}|$  désigne la cardinalité d'un ensemble non vide  $\mathbf{u} \subset \{1, \dots, d\}$ , par exemple, pour  $\mathbf{u} = \{u_1, \dots, u_\ell\} \subset \{1, \dots, d\}$ ,  $|\mathbf{u}| = \ell$ ;
- iv) Pour  $x \in \mathcal{X}$  et  $\mathbf{u} \subset \{1, \dots, d\}$ , nous désignons par  $x^{(\mathbf{u})}$  le point (vecteur) extrait de  $x = (x^{(1)}, \dots, x^{(d)})' \in \mathcal{X}$  dont les composantes sont indexées par les indices de  $\mathbf{u}$ ,  $x^{(\mathbf{u})} = (x^{(u_1)}, \dots, x^{(u_\ell)})'$ .

Dans la notation (iv),  $x'$  désigne la transposée du vecteur  $x$ .

Pour définir l'inégalité de Koksma-Hlwaka généralisée, l'hypothèse suivante doit être vérifiée (voir Hickernell (1998)), pour une fonction  $f(\cdot)$  donnée, définie sur  $\mathcal{X} = [0, 1]^d$ .

**HYPOTHÈSE 1.** — *Toute dérivée partielle croisée d'ordre inférieur ou égal à  $d$  de  $f(\cdot)$  est intégrable d'ordre  $p$ , où  $1 \leq p < \infty$  est un entier :*

$$f(\cdot) \in \mathcal{W}_p(\mathcal{X}) \equiv \left\{ \begin{array}{l} f := \frac{\partial^{|\mathbf{u}|} f}{\partial x^{(\mathbf{u})}} = \frac{\partial^{|\mathbf{u}|} f}{\partial x^{(u_1)} \dots \partial x^{(u_\ell)}} \in L^p([0, 1]^{|\mathbf{u}|}), \\ \forall \mathbf{u} = \{u_1, \dots, u_\ell\} \subseteq \{1, \dots, d\} \end{array} \right\}. \quad (4)$$

### 2.2. Discrétance et inégalité généralisée de Koksma-Hlwaka

De façon générale, sous l'hypothèse 1, pour une suite de points  $\mathbf{x}(n) = \{x_1, \dots, x_n\}$  dans  $\mathcal{X}$ , l'inégalité de Koksma-Hlwaka généralisée peut s'écrire comme suit (voir, Niederreiter (1998)) :

$$|I(f) - \hat{I}_{\mathbf{x}(n)}(f)| = \left| \int_{\mathcal{X}} f(x) d(F_n(x) - U(x)) \right| \leq V(f) D(\mathbf{x}(n)), \quad (5)$$

où :

- i)  $I(f) = \int_{\mathcal{X}} f(x)dx$ , où  $dx = dU(x)$  est la mesure de Lebesgue sur  $\mathcal{X} = [0, 1]^d$ ,
- ii)  $\hat{I}_{\mathbf{x}(n)}(f) = \frac{1}{n} \sum_{x_i \in \mathbf{x}(n)} f(x_i)$  avec  $\mathbf{x}(n) = \{x_1, \dots, x_n\}$ ,
- iii)  $F_n$  désigne la fonction de répartition empirique de l'échantillon  $\mathbf{x}(n) = \{x_1, \dots, x_n\}$  dans  $[0, 1]^d$ ,
- iv)  $U$  désigne la fonction de répartition de la loi uniforme sur  $[0, 1]^d$ ,
- v)  $V(f)$  est une *variation* de  $f(\cdot)$  (voir plus loin),
- vi) et  $D(\mathbf{x}(n))$  est un terme qui dépend des points  $\mathbf{x}(n) = \{x_1, \dots, x_n\}$ , et correspond à la notion de *discrédance* (voir plus loin).

• Dans le cas de l'inégalité classique de Koksma-Hlwaka (voir Hlwaka (1961) et Niederreiter (1998)),  $V(f)$  désigne la *variation totale au sens de Hardy et Krause* de  $f(\cdot)$ , et  $D(\mathbf{x}(n))$  la *discrédance à l'origine* de  $\mathbf{x}(n)$ , usuellement notée  $D^*$ , et définie par :

$$D^*(\mathbf{x}(n)) := \|F_n - U\|_{L^\infty(\mathcal{X})}. \quad (6)$$

Dans le cas où la suite  $\mathbf{x}(n) = \{x_1, \dots, x_n\}$  est la réalisation d'une suite de variables aléatoires indépendantes identiquement distribuées,  $D^*(\mathbf{x}(n))$  correspond à la statistique de Kolmogorov-Smirnov (voir les inégalités de Dvoretzky *et al.*, et de Massart (1990) lorsque  $d = 1$ , pour l'étude de cette statistique).

• Dans le cas de l'inégalité de Koksma-Hlwaka généralisée, la définition de la discrédance  $D(\mathbf{x}(n))$  fait usage d'une norme convenable de l'application  $(I - \hat{I}_{\mathbf{x}(n)})$  définie par :

$$f \rightarrow \text{Err}(f, \mathbf{x}(n)) = (I - \hat{I}_{\mathbf{x}(n)})(f). \quad (7)$$

La définition de la variation  $V(f)$  utilise une norme de la fonction  $f(\cdot)$  dans un espace de Hilbert à noyau auto-reproduisant (en abrégé, RKHS pour *Reproducing Kernel Hilbert Space*, voir Hickernell (1998)). Comme il est possible de définir différents noyaux et, par conséquent, différentes normes sur de tels espaces, on obtient de nombreuses variantes de ce résultat.

Soit un espace de Hilbert  $\mathcal{H} \equiv \mathcal{W}_2(\mathcal{X})$  de fonctions sur  $\mathcal{X}$  (selon la notation (4) de l'Hypothèse 1, prise pour  $p = 2$ ) muni d'un noyau auto-reproduisant  $K$ , symétrique, de type positif, et de carré intégrable.

$$K(x, y) = K(y, x), \quad \forall x, y \in [0, 1]^d \quad (8)$$

$$\sum_{i, k} a_i a_k K(x_i, x_k) \geq 0, \quad \forall a_i \in \mathbb{R}, \quad x_i \in [0, 1]^d \quad (9)$$

$$\int_{[0, 1]^d} K(x, x) dx < \infty. \quad (10)$$

Soit  $f \in \mathcal{H}$  une fonction sur  $\mathcal{X}$ . Lorsque  $f(\cdot)$  est constante, nous avons  $\text{Err}(f, \mathbf{x}(n)) = 0$  dans (7). Nous désignons par  $f_{\perp}$  la projection de  $f(\cdot)$  sur le sous-espace de  $\mathcal{X}$  orthogonal à la fonction 1 (le fait que  $1 \in \mathcal{H}$  est conséquence de (10)). On notera  $\langle \cdot, \cdot \rangle_K$  le produit scalaire induit par le noyau  $K$ , et  $\|\cdot\|_K = (\langle \cdot, \cdot \rangle_K)^{1/2}$ , la norme correspondante. Nous avons  $\text{Err}(f, \mathbf{x}(n)) := \text{Err}(f_{\perp}, \mathbf{x}(n))$ ,  $\forall f(\cdot) \in \mathcal{H}$ . Par le théorème de représentation de Riesz (voir Riesz et Nagy (1955)), il existe  $\xi \in \mathcal{H}$  tel que :  $\text{Err}(f, \mathbf{x}(n)) = \langle \xi, f \rangle_K$ ,  $\forall f \in \mathcal{H}$ . Ainsi :

$$\text{Err}(f, \mathbf{x}(n)) = \text{Err}(f_{\perp}, \mathbf{x}(n)) = \langle \xi, f_{\perp} \rangle_K .$$

Par application de l'inégalité de Cauchy-Schwarz, on obtient :

$$\text{Err}(f, \mathbf{x}(n)) \leq \|f_{\perp}(\cdot)\|_K \|\xi\|_K . \quad (11)$$

L'inégalité (11) est appelée *inégalité de Koksma-Hlwaka généralisée* (Niederreiter (1998)). La *discrédance généralisée* de  $\mathbf{x}(n)$  est alors définie par  $D(\mathbf{x}(n)) = D^{L^2}(\mathbf{x}(n), K) = \|\xi\|_K$  (voir Hickernell (1998)). En utilisant la forme explicite de  $\xi$ , la *discrédance généralisée* ci-dessus s'exprime comme suit :

$$D^{L^2}(\mathbf{x}(n), K) := \left\{ \int_{\mathcal{X}^{2d}} K(x, y) d(F_n(x) - U(x)) d(F_n(y) - U(y)) \right\} . \quad (12)$$

Dans Hickernell (1998), différents noyaux sont proposés faisant intervenir des polynômes de Bernouilli. Les discrédances  $L^2$  *modifiée*  $DM^{L^2}(\mathbf{x}(n))$ ,  $L^2$  *centrée*  $DC^{L^2}(\mathbf{x}(n))$ , et  $L^2$  *symétrique*,  $DS^{L^2}(\mathbf{x}(n))$  y sont notamment définies. L'intérêt pratique de ces trois discrédances est qu'elles sont facilement calculables quelle que soit la dimension  $d$  de l'espace  $\mathcal{X} = [0, 1]^d$ . De simples formules analytiques existent à cet effet. Ceci n'est pas le cas de la *discrédance à l'origine*, définie en (6),  $D^*(\mathbf{x}(n))$ , dont le calcul est délicat en dimension  $d > 2$  (voir Thiémarc (2000)). De plus, les discrédances  $L^2$  *modifiée*,  $L^2$  *centrée*, et  $L^2$  *symétrique* ont une interprétation géométrique simple. Ce sont des comparaisons entre la proportion de points de  $\mathbf{x}(n)$  compris dans des pavés de  $[0, 1]^d$  (la mesure empirique) et le volume de ces pavés. Elles définissent donc bien des caractérisations de la disposition des points  $\mathbf{x}(n) = \{x_1, \dots, x_n\}$  dans l'espace  $\mathcal{X} = [0, 1]^d$ .

Précisons cependant que l'utilisation de ce critère n'est théoriquement valable que lorsqu'on considère certaines classes de fonctions. En effet, par sa définition, la discrédance fait intervenir le noyau d'un espace fonctionnel de Hilbert. Pour que  $D^{L^2}(\mathbf{x}(n), K)$  soit bornée, ce noyau doit être de carré intégrable. Ceci implique notamment que les fonctions à considérer doivent être bornées. C'est l'expression du noyau qui permet de caractériser plus précisément les différentes propriétés de la classe de fonctions. La seule propriété d'intégrabilité n'est pas suffisante et ces fonctions possèdent le plus souvent des propriétés de régularité « fortes » (comme celles vérifiant l'Hypothèse 1). A notre connaissance, l'étude de la discrédance et de son

calcul a été essentiellement abordé en considérant des RKHS de noyaux faisant intervenir des polynômes de Bernoulli (*cf.* Hickernell (1998)). Une perspective de recherche consisterait en l'étude d'autres noyaux. Ceux-ci permettraient peut-être de «relâcher» l'Hypothèse 1 (ordre d'intégrabilité inférieur, par exemple).

Il est aussi possible de définir des inégalités de Koksma-Hlwaka analogues à (11) pour des espaces de Banach  $\mathcal{W}_p(\mathcal{X})$  (avec  $p \neq 2$  dans (4) de l'Hypothèse 1). Il suffit alors de remplacer l'inégalité (11) de Cauchy-Schwarz par une inégalité de Hölder.

### 3. Majoration de critères

#### 3.1. Introduction

Rappelons que nous nous plaçons dans le contexte de l'estimation  $\hat{f}_n(\cdot)$  d'un paramètre fonctionnel  $f(\cdot)$  par la méthode des fonctions orthogonales décrite par (1), (2) et (3). Dans ce qui suit, nous nous intéresserons à la majoration de critères exprimant la qualité de l'estimation  $\hat{f}_n(\cdot)$  de  $f(\cdot)$ , l'IMSE, et la MSE, définis comme suit :

$$\text{IMSE}(f, \hat{f}_n) := \int_{\mathcal{X}} \mathbb{E}(\hat{f}_n(x) - f(x))^2 dx, \quad \text{et} \quad \text{MSE}(\hat{a}_1, a_1) := \mathbb{E}(\hat{a}_1 - a_1)^2. \quad (13)$$

Notons que lorsque la première fonction du système orthogonal  $\{v_k(\cdot) : k \geq 1\}$  est égale à 1,  $v_1 = 1$ ,  $\text{MSE}(\hat{a}_1, a_1)$  correspond à un critère de robustesse d'estimation de la moyenne  $a_1 = \int_{\mathcal{X}} f(x) dx$  (voir (2) et (3)).

Par la suite, le nombre  $N$  des fonctions orthogonales  $\{v_1(\cdot), \dots, v_N(\cdot)\} \subset L^2(\mathcal{X})$  utilisées pour l'estimation de  $f(\cdot)$  (*cf.* (2) et (3)) sera alors supposé spécifié (voir par exemple Bosq et Bluez (1978), Aubin (2005) ou Devroye et Györfi (1985)).

Les majorations de l'IMSE et de la MSE que nous proposerons seront obtenues à l'aide de l'inégalité, (11) de Koksma-Hlwaka généralisée, appliquée aux fonctions  $f(\cdot)$  et  $v_k(\cdot)$ ,  $1 \leq k \leq N$ . Aussi, nous supposerons que  $f(\cdot)$  vérifie l'Hypothèse 1 présentée au §2.1 avec  $p = 2$  dans (4),  $f \in \mathcal{C}(\mathcal{X}) \cap \mathcal{W}_2(\mathcal{X})$ . Pour la majoration de l'IMSE, nous ferons l'hypothèse que les fonctions  $\{v_k(\cdot) : 1 \leq k \leq N\}$  de  $L^2(\mathcal{X})$  forment un système orthonormal de  $L^2(\mathcal{X})$ , qu'elles sont continues et vérifient l'hypothèse 1.

L'application de l'inégalité (11) de Koksma-Hlwaka permettra de faire apparaître les discrécances généralisées de type  $L^2$  (*cf.* (12)) dans les termes de majoration de l'IMSE et de la MSE.

#### 3.2. Majoration de l'IMSE

A l'aide de l'inégalité classique de Koksma-Hlwaka (*cf.* (5) et (6)), il est possible de majorer l'IMSE par un terme dépendant de la *discrécance* à l'origine (on se référera aux arguments de Rafajlowicz et Schwabe (2006)).

En reprenant cette approche, et en la généralisant au cas des discrédances généralisées de type  $L^2$ , nous avons obtenu la majoration suivante :

$$\begin{aligned} \text{IMSE}(f, \hat{f}_n) &\leq \frac{N\sigma^2}{n} \left[ 1 + 2 M_v C_{V(v)} D^{L^2}(\mathbf{x}(n), K) \right] \\ &\quad + N [C_{V(f)} M_v + M_f C_{V(v)}]^2 D^{L^2}(\mathbf{x}(n), K)^2 \\ &\quad + R(N, f). \end{aligned} \quad (14)$$

Ici,  $C_{V(v)}$ , et  $C_{V(f)}$ , sont des constantes qui dépendent, respectivement, des fonctions  $v_k(\cdot)$ ,  $k = 1, \dots, N$ , et  $f(\cdot)$ .  $M_v$ , et  $M(f)$ , sont des constantes qui majorent, respectivement, les normes uniformes, des  $v_k(\cdot)$ , pour  $k = 1, \dots, N$ , et de la fonction  $f(\cdot)$ . Le terme  $R(N, f)$  est un terme d'erreur. Enfin,  $D^{L^2}(\mathbf{x}(n), K)$  désigne une discrédance généralisée de type  $L^2$  (voir équation (12)).

**Démonstration de l'inégalité (14)** (inspirée de Rafajłowicz et Schwabe (2006))

A) *Rappel des Hypothèses et notations supplémentaires*

Par la suite, nous ne traiterons que le cas où l'on considère le RKHS  $\mathcal{H} \equiv \mathcal{W}_2(\mathcal{X})$  permettant de définir la discrédance  $L^2$  modifiée  $\text{DM}^{L^2}(\mathbf{x}(n))$  (voir Hickernell (1998) pour sa définition précise). L'application de l'inégalité (11) de Koksma-Hlwaka à une fonction  $g \in \mathcal{H}$  en un ensemble de points  $\mathbf{x}(n) = \{x_1, \dots, x_n\} \in \mathcal{X}^n$  fera donc intervenir  $\text{DM}^{L^2}(\mathbf{x}(n))$  (cf. §2.2). Les cas des RKHS  $\mathcal{H}$  muni des noyaux permettant d'obtenir les discrédances  $L^2$  centrée et  $L^2$  symétrique étant similaire, le détail des calculs ne sera pas précisé. Nous noterons :

- i)  $\partial^{|\mathbf{u}|} g / \partial x^{(\mathbf{u})}$ , la dérivée partielle croisée d'ordre  $|\mathbf{u}|$  par rapport à  $x^{(\mathbf{u})} = (x^{(u_1)}, \dots, x^{(u_\ell)})$ , (voir (4) et les notations (i), (ii), (iii) (iv) du paragraphe 2);
- ii)  $x^{(\mathbf{u}^c)} = (1, \dots, 1)$ , l'ensemble des points  $x = (x(1), \dots, x^{(d)}) \in \mathcal{X} = [0, 1]^d$  tels que  $x^{(m)} = 1$ , pour tout  $m \notin \mathbf{u} = \{u_1, \dots, u_\ell\} \subseteq \{1, \dots, d\}$ . De manière explicite,
 
$$\{x^{(\mathbf{u}^c)} = (1, \dots, 1)\} := \{x : x \in \mathcal{X} = [0, 1]^d \text{ tel que } : x^{(j)} = 1, \forall j \in \{1, \dots, d\}, \text{ et } j \notin \mathbf{u}\};$$
- iii)  $g|_{\{x^{(\mathbf{u}^c)} = (1, \dots, 1)\}}$ , la restriction d'une fonction  $g$  sur  $\{x^{(\mathbf{u}^c)} = (1, \dots, 1)\}$ ;
- iv) pour une fonction  $g(\cdot) \in \mathcal{W}_2(\mathcal{X})$  et pour un ensemble non vide d'indices  $\mathbf{u} \subset \{1, \dots, d\}$ ,

$$VM^{(\mathbf{u})}(g) := \int_{[0,1]^{|\mathbf{u}|}} \left[ \frac{\partial^{|\mathbf{u}|}(g(x))}{\partial x^{(\mathbf{u})}} \Big|_{x^{(\mathbf{u}^c)} = (1, \dots, 1)} \right] dx^{(\mathbf{u})}. \quad (15)$$

Nous désignerons la norme de la fonction  $g$  orthogonale à 1 dans  $\mathcal{H}$  par

$$\mathcal{V}_{\mathcal{M}}(g) := \|g(\cdot)_{\perp}\|_K, \quad (16)$$

sous le nom de *variation modifiée*. Pour  $g \in \mathcal{H}$ , il ressort de Hickernell (1998) que

$$\mathcal{V}_{\mathcal{M}}(g) := \left[ \sum_{\mathbf{u} \subset \{1, \dots, d\}} \left\| \frac{\partial^{|\mathbf{u}|} g}{\partial \mathbf{x}^{(\mathbf{u})}} \Big|_{\{x^{(u^c)} = (1, \dots, 1)\}} \right\|_{L^2([0,1]^{|\mathbf{u}|})}^2 \right]^{1/2}, \quad (17)$$

$$:= \left[ \sum_{\mathbf{u} \subset \{1, \dots, d\}} \left( VM^{(\mathbf{u})}(g) \right)^2 \right]^{1/2}, \quad \text{selon nos notations.} \quad (18)$$

### B) Décomposition de l'IMSE

En utilisant le théorème de Fubini pour  $f \in L^2(\mathcal{X})$ , et  $v_k(\cdot) \in L^2(\mathcal{X})$ , et par orthogonalité et normalité des  $\{v_k, k \geq 1\}$ , nous pouvons obtenir,

$$\begin{aligned} \text{IMSE}(f, \hat{f}_n) &= \sum_{k=1}^N \text{Var}(\hat{a}_k) + \sum_{k=1}^N [\mathbb{E}(\hat{a}_k - a_k)]^2 + \sum_{k \geq N+1} a_k^2 \\ &= W_n + B_n^2 + R(N, f), \end{aligned}$$

avec

$$W_n := \sum_{k=1}^N \text{Var}\{\hat{a}_k\}, \quad B_n^2 := \sum_{k=1}^N (\mathbb{E}\{\hat{a}_k - a_k\})^2, \quad R(N, f) := \sum_{k \geq N+1} a_k^2. \quad (19)$$

Nous allons, à présent, majorer le terme de variance,  $W_n$ , et le terme de biais,  $B_n^2$ , à l'aide de l'inégalité de Koksma-Hlwaka généralisée (voir §2).

### C) Majoration du terme de Variance $W_n$

Considérons la composante  $\text{Var}\{\hat{a}_k\}$  de  $W_n$  dans (19). Par hypothèse d'indépendance des  $y_i$  (voir (1)), nous avons :

$$\begin{aligned} \text{Var}\{\hat{a}_k\} &= \text{Var} \left\{ \frac{1}{n} \sum_{i=1}^n v_k(x_i) y_i \right\} = \frac{1}{n^2} \sum_{i=1}^n v_k(x_i)^2 \text{Var}(y_i) \\ &= \frac{\sigma^2}{n} \left\{ \frac{1}{n} \sum_{i=1}^n v_k(x_i)^2 \right\}. \end{aligned} \quad (20)$$



• D'après les Hypothèses du §3.1, nous pouvons appliquer au membre de droite de (20) l'inégalité (11) de Koksma-Hlwaka généralisée,

$$\left| \frac{1}{n} \sum_{i=1}^n v_k(x_i)^2 - \int_{\mathcal{X}} v_k(x)^2 dx \right| \leq \mathcal{V}_{\mathcal{M}}(v_k^2) \text{DM}^{L^2}(\mathbf{x}(n)). \quad (21)$$

L'inégalité précédente implique que

$$\left| \frac{1}{n} \sum_{i=1}^n v_k(x_i)^2 \right| \leq \left| \int_{\mathcal{X}} v_k(x)^2 dx \right| + \mathcal{V}_{\mathcal{M}}(v_k^2) \text{DM}^{L^2}(\mathbf{x}(n)). \quad (22)$$

Nous posons maintenant

$$\left( VM_2^{(u)}(v_k^2) \right)^2 := \int_{[0,1]^{|u|}} \left[ \frac{\partial^{|u|}(v_k^2(x))}{\partial x^{(u)}} \Big|_{x^{(u^c)}=(1,\dots,1)} \right]^2 dx^{(u)}. \quad (23)$$

Par différentiation du terme  $v_k^2(x)$  dans l'intégrale (23), nous obtenons :

$$\left( VM^{(u)}(v_k^2) \right)^2 = 4 \int_{[0,1]^{|u|}} \left[ \left( \frac{\partial^{|u|}(v_k(x))}{\partial x^{(u)}} v_k(x) \right) \Big|_{x^{(u^c)}=(1,\dots,1)} \right]^2 dx^{(u)}. \quad (24)$$

Par hypothèses sur les  $v_k(\cdot)$ , nous pouvons appliquer l'inégalité de Hölder à (24) pour obtenir les inégalités,

$$\begin{aligned} \left( VM^{(u)}(v_k^2) \right)^2 &\leq 4 \|v_k^2\|_{L^\infty([0,1]^{|u|})} \left\| \left( \frac{\partial^{|u|}(v_k)}{\partial x^{(u)}} \Big|_{x^{(u^c)}=(1,\dots,1)} \right)^2 \right\|_{L^1([0,1]^{|u|})} \\ &\leq 4 \sup_{x \in \mathcal{X}} |v_k(x)^2| \int_{[0,1]^{|u|}} \left( \frac{\partial^{|u|}(v_k)}{\partial x^{(u)}} \Big|_{x^{(u^c)}=(1,\dots,1)} \right)^2 dx^{(u)} \\ &\leq 4 \sup_{x \in \mathcal{X}} |v_k(x)^2| \left( VM^{(u)}(v_k) \right)^2. \end{aligned}$$

La *variation modifiée* d'une fonction étant définie, comme en (17) et (18), en sommant les termes de gauche et droite de cette inégalité sur tous les sous-ensembles d'indices  $\mathbf{u} = \{u_1, \dots, u_\ell\}$  non vides possibles de  $\{1, \dots, d\}$ , on a, par continuité des  $v_k(\cdot)$ ,

$$\begin{aligned} \left[ \sum_{\mathbf{u} \subset \{1, \dots, d\}} VM^{(u)}(v_k^2) \right]^{1/2} &= \mathcal{V}_{\mathcal{M}}(v_k^2) \\ &\leq 2 \sup_{x \in \mathcal{X}} |v_k(x)| \mathcal{V}_{\mathcal{M}}(v_k). \end{aligned} \quad (25)$$

Dans le cadre de l'estimation par projection induite par les équations (1), (2) et (3), nous considérons les  $N$  premières fonctions  $v_1(\cdot), \dots, v_N(\cdot)$  dans  $L^2(\mathcal{X})$ , où  $N \geq 1$  est un entier spécifié. Ainsi, pour  $N$  fixé, posons

$$C_{\mathcal{V}_{\mathcal{M}}(v), N} := \max \{ \mathcal{V}_{\mathcal{M}}(v_1), \dots, \mathcal{V}_{\mathcal{M}}(v_N) \},$$

$$M_v := \max \left\{ \sup_{x \in \mathcal{X}} |v_1(x)|, \dots, \sup_{x \in \mathcal{X}} |v_N(x)| \right\}.$$

Avec ces notations, et en utilisant l'Hypothèse d'orthonormalité des  $v_k(\cdot)$ , pour  $k = 1, \dots, N$ , les inégalités (22) et (25) impliquent que

$$\left| \frac{1}{n} \sum_{i=1}^n v_k(x_i)^2 \right| \leq 1 + 2 M_v C_{\mathcal{V}_{\mathcal{M}}(v), N} \text{DM}^{L^2}(\mathbf{x}(n)). \quad (26)$$

Les relations (20) et (26) impliquent que

$$\text{Var}(\hat{a}_k) \leq \frac{\sigma^2}{n} \left[ 1 + 2 M_v C_{\mathcal{V}_{\mathcal{M}}(v), N} \text{DM}^{L^2}(\mathbf{x}(n)) \right]. \quad (27)$$

Par sommation sur  $k$  de 1 à  $N$ , le terme de variance  $W_n$  dans (19) est majoré par

$$W_n = \sum_{k=1}^N \text{Var}(\hat{a}_k) \leq \frac{N \sigma^2}{n} \left[ 1 + 2 M_v C_{\mathcal{V}_{\mathcal{M}}(v), N} \text{DM}^{L^2}(\mathbf{x}(n)) \right]. \quad (28)$$

• De la même façon, en considérant les RKHS permettant d'obtenir les discrédances  $L^2$  modifiée  $\text{DM}^{L^2}(\mathbf{x}(n))$ ,  $L^2$  centrée  $\text{DC}^{L^2}(\mathbf{x}(n))$ , et  $L^2$  symétrique,  $\text{DS}^{L^2}(\mathbf{x}(n))$  (voir Hickernell (1998)) il est possible d'obtenir les inégalités

$$W_n \leq \frac{N \sigma^2}{n} \left[ 1 + 2 M_v C_{\mathcal{V}_c(v), N} \text{DC}^{L^2}(v_k) \right], \quad (29)$$

$$W_n \leq \frac{N \sigma^2}{n} \left[ 1 + 2 M_v C_{\mathcal{V}_s(v), N} \text{DS}^{L^2}(v_k) \right], \quad (30)$$

où  $\text{DC}^{L^2}(\mathbf{x}(n))$  désigne la discrédance  $L^2$  centrée,  $\text{DS}^{L^2}(\mathbf{x}(n))$ , la discrédance  $L^2$  symétrique, voir le §2.1, et  $C_{\mathcal{V}_c(v), N}$  et  $C_{\mathcal{V}_s(v), N}$ , des constantes conve-nables.

#### D) Majoration du terme de biais

Par définition (voir les équations (1), (2) et (3)), nous avons, pour tout  $k \in \{1 \dots N\}$ ,

$$\mathbb{E}(\hat{a}_k) - a_k = \frac{1}{n} \sum_{i=1}^n f(x_i) v_k(x_i) - \int_{\mathcal{X}} f(x) v_k(x) dx. \quad (31)$$

D'après les hypothèses du §3.1, l'inégalité (11) de Koksma-Hlwaka s'applique. Conformément aux notations du §2.2, nous déduisons de celle-ci et de (31) que, pour  $k \in \{1, \dots, N\}$ ,

$$|\mathbb{E}(\hat{a}_k) - a_k| \leq \mathcal{V}_{\mathcal{M}}(f(\cdot)v_k(\cdot)) \text{DM}^{L^2}(\mathbf{x}(n)). \quad (32)$$

Nous effectuons un raisonnement analogue à celui utilisé plus haut lors de la majoration du terme de variance  $W_n$  (inégalités (28), (29), (30)). La modification consiste à remplacer les fonctions  $v_k^2(\cdot)$  de (21) par les fonctions  $f(\cdot) \times v_k(\cdot)$  de (32). Par application de (11), nous obtenons

$$B_n^2 \leq N [C_{\mathcal{V}_{\mathcal{M}}(f)} M_v + M_f C_{\mathcal{V}_{\mathcal{M}}(v), N}]^2 \text{DM}^{L^2}(\mathbf{x}(n))^2, \quad (33)$$

$$B_n^2 \leq N [C_{\mathcal{V}_c(f)} M_v + M_f C_{\mathcal{V}_c(v), N}]^2 \text{DC}^{L^2}(\mathbf{x}(n))^2, \quad (34)$$

$$B_n^2 \leq N [C_{\mathcal{V}_s(f)} M_v + M_f C_{\mathcal{V}_s(v), N}]^2 \text{DS}^{L^2}(\mathbf{x}(n))^2. \quad (35)$$

### E) Majoration de l'IMSE

D'après (19), par sommation des expressions obtenues ci-dessus en (28), (29), (30), et, en (33), (34), (35), respectivement, comme majorations du terme de variance  $W_n$ , et de biais,  $B_n^2$ , nous aboutissons à l'inégalité (14), soit

$$\begin{aligned} \text{IMSE}(f, \hat{f}_n) &\leq \frac{N\sigma^2}{n} \left[ 1 + 2 M_v C_{V(v)} \text{D}^{L^2}(\mathbf{x}(n), K) \right] \\ &\quad + N [C_{V(f)} M_v + M_f C_{V(v)}]^2 \text{D}^{L^2}(\mathbf{x}(n), K)^2 \\ &\quad + R(N, f). \end{aligned} \quad (36)$$

Remarquons que l'inégalité (32) permet d'obtenir une majoration de la MSE de toutes les estimations,  $\hat{a}_k$   $k \in \{1, \dots, N\}$ , des coefficients issus de la régression par projection orthogonale (*cf.* (2) et (3)). En effet, par définition de la MSE (voir (13)), nous avons,

$$\begin{aligned} \text{MSE}(\hat{a}_k, a_k) &= \mathbb{E}[\hat{a}_k - \mathbb{E}(a_k)]^2 + [\mathbb{E}(a_k) - a_k]^2 \\ &\leq \text{Var}(\hat{a}_k) + \left[ \mathcal{V}_{\mathcal{M}}(f(\cdot)v_k(\cdot)) \text{DM}^{L^2}(\mathbf{x}(n)) \right]^2. \end{aligned} \quad (37)$$

### 3.3. Majoration de la MSE

Faisons l'hypothèse que la première fonction du système  $\{v_1(\cdot), \dots, v_N(\cdot)\} \subset L^2(\mathcal{X})$  utilisé dans (2) et (3) est la constante 1,  $v_1 = 1$  (ce qui est rendu possible par (10)), Supposons aussi que la fonction  $f(\cdot)$  vérifie l'Hypothèse 1 avec  $p = 2$  dans (4). En faisant usage, soit d'une approche bayésienne, soit de l'utilisation de l'inégalité de Koksma-Hlwaka généralisée (voir (37)), il est possible de majorer le critère, dit de *robustesse*,  $\text{MSE}(\hat{a}_1, a_1)$  en (13), par

$$\text{MSE}(\hat{a}_1, a_1) \leq \frac{\sigma^2}{n} + \left[ \text{D}^{L^2}(\mathbf{x}(n), K) V(f) \right]^2, \quad (38)$$

où  $V(f)$  désigne une *variation de  $f(\cdot)$*  et  $D^{L^2}(\mathbf{x}(n), K)$  une *discrépance généralisée* de type  $L^2$  (voir §2.2 équation (12)). Nous renvoyons à Hickernell (1999) pour une démonstration précise de cette inégalité.

### 3.4. Interprétation

Les inégalités (14) et (38) montrent que les termes dominant les critères d'IMSE et de MSE, font intervenir la discrépance généralisée. Ces majorations montrent que la caractérisation de la disposition des points dans l'espace  $\mathcal{X} = [0, 1]^d$ , exprimée par la discrépance généralisée, influe sur la qualité de l'estimation de  $f(\cdot)$ .

Pour pouvoir utiliser au mieux cette propriété selon les données disponibles, il est nécessaire de définir un cadre méthodologique, ce qui est fait dans le §4 ci-dessous.

## 4. Cadre méthodologique

L'objectif de ce paragraphe sera la *construction*, la *sélection*, ou la *spécification* d'un ensemble de points  $\mathbf{x}(n) \in \mathcal{X}^n$  apte à fournir une estimation  $\hat{f}_n(\cdot)$  satisfaisante (au sens de l'IMSE et de la MSE) de  $f(\cdot)$ .

Les inégalités (14) et (38) montrent que, sous certaines hypothèses, l'IMSE et la MSE sont majorés par un terme faisant intervenir la discrépance de  $\mathbf{x}(n)$  noté  $D^{L^2}(\mathbf{x}(n), K)$  (voir la définition (12) §2.2). Plus précisément, leur étude permet de formuler les remarques suivantes.

*Remarque 1.* — Lorsque  $n$  augmente et  $D^{L^2}(\mathbf{x}(n), K)$  diminue, l'IMSE et que la MSE diminuent. Signalons de plus que lorsque  $n \rightarrow \infty$ , sous certaines hypothèses supplémentaires, sur la fonction  $f(\cdot)$ , les fonctions  $v_k(\cdot)$  et l'ordre de troncature  $N$  en fonction de  $n$ , l'utilisation de *suites à discrépance faible* assure une vitesse de convergence optimale des estimations  $\hat{f}_n(\cdot)$  vers  $f(\cdot)$  au sens de l'IMSE (Stone-1982 et Rafałłowicz-2005).

*Remarque 2.* — Lorsque  $n$  est constant et que  $D^{L^2}(\mathbf{x}(n), K)$  diminue, l'IMSE et la MSE diminuent.

*Remarque 3.* — Lorsque  $n$  diminue et  $D^{L^2}(\mathbf{x}(n), K)$  diminue, l'étude des inégalités (14) et (38) ne permet pas de connaître l'évolution de la MSE et de l'IMSE. En effet, elles font toutes deux intervenir un terme en  $1/n$  qui augmente dans ce contexte.

Le cadre méthodologique défini ci-dessous permet de prendre en compte ces remarques de façon pratique. Nous distinguerons différents cas de figures en fonction des données (« points d'observation », « observations ») disponibles, c'est-à-dire des  $(x_i, y_i)$ ,  $i = 1, \dots, n$  selon nos notations, cf. (1)).

**Cas 1.** *Absence initiale de la base de données  $\mathbf{x}(n) = \{x_1, \dots, x_n\}$  dans  $[0, 1]^d$*

Compte tenu de la Remarque 1, lorsqu'on peut maîtriser le choix des points  $\mathbf{x}(n)$  dans  $\mathcal{X} = [0, 1]^d$ , il est préférable de choisir une suite à *discrépance faible*.

**Cas 2.** *Choix d'un nombre « imposé » de points dans une base de données  $\mathbf{x}(n) = \{x_1, \dots, x_n\}$*

Lorsqu'on doit choisir un ensemble  $\mathbf{x}_1(n_1)$  de  $n_1$  points dans un ensemble de points candidats  $\mathbf{x}(n) = \{x_1, \dots, x_n\}$ , on pourra commencer par étudier la qualité de la base  $\mathbf{x}(n)$  à l'aide de critères faisant intervenir la discrépance (voir le §5.2). Compte tenu de la Remarque 2, on sélectionnera l'ensemble de points dont la quantité  $D^{L^2}(\mathbf{x}_1(n_1))$  est la plus faible. Là encore, une étude de l'ensemble des points sélectionnés sera menée (cf. §5.2).

**Cas 3.** *Choix « libre » de points dans une base de données  $\mathbf{x}(n) = \{x_1, \dots, x_n\}$*

Nous devons ici choisir un ensemble de points  $\mathbf{x}_1(n_1)$  dans un ensemble  $\mathbf{x}(n) = \{x_1, \dots, x_n\} \in \mathcal{X}^n$  ( $n_1$  n'est pas imposé). Compte tenu de la Remarque 3, ce choix est délicat. Nous proposons la sélection de points  $\mathbf{x}_1(n_1)$  parmi  $\mathbf{x}(n)$  qui permettent de réduire la quantité  $Q(\mathbf{x}(n)) := D^{L^2}(\mathbf{x}(n), K)/n$ . Le nombre  $n_1$  de points sélectionnés sera plus important que si l'on cherche simplement à diminuer la quantité  $D^{L^2}(\mathbf{x}(n), K)$ . Ainsi, on limitera l'augmentation du terme en  $1/n$  dans les majorations de l'IMSE et de la MSE décrites par les inégalités (14) et (38). De plus, la diminution de  $Q(\mathbf{x}(n))$  implique la diminution de  $D^{L^2}(\mathbf{x}(n), K)$ . En effet, si

$$\frac{D^{L^2}(\mathbf{x}(n-1), K)}{n-1} \leq \frac{D^{L^2}(\mathbf{x}(n), K)}{n},$$

ce qui implique,

$$D^{L^2}(\mathbf{x}(n-1), K) \leq D^{L^2}(\mathbf{x}(n), K)^2. \quad (39)$$

Bien entendu, une étude préalable de la qualité de  $\mathbf{x}(n)$ , puis de  $\mathbf{x}_1(n_1)$  à l'aide de critères faisant intervenir la discrépance sera menée (voir §5.2).

**Cas 4** *Données  $(x_i, y_i)$ ,  $i = 1, \dots, n$  disponibles*

Il est d'usage d'appeler les couples  $(x_i, y_i)$ ,  $i = 1, \dots, n$  des données (point d'observation, observation). Tout d'abord une estimation  $\hat{f}_n(\cdot)$  de  $f(\cdot)$  à l'aide des données initiales  $(x_i, y_i)$ ,  $i = 1, \dots, n$ , sera réalisée conformément à (3). Parmi les données  $(x_i, y_i)$ ,  $i = 1, \dots, n$ , nous chercherons un sous-ensemble de nature à fournir une estimation satisfaisante  $\hat{f}_{n_1}(\cdot)$  de  $f(\cdot)$  (au sens de la MSE et de l'IMSE). Dans ce contexte, nous sélectionnerons un sous-ensemble  $\mathbf{x}_1(n_1) \subset \mathbf{x}(n) = \{x_1, \dots, x_n\}$  de la même façon que celle décrite au cas 3. L'estimation  $\hat{f}_{n_1}(\cdot)$  sera réalisée à l'aide des points de  $\mathbf{x}_1(n)$  et des observations correspondantes (selon (3)). On comparera les estimations  $\hat{f}_n(\cdot)$

et  $\hat{f}_{n_1}(\cdot)$  de  $f(\cdot)$ . Ceci se fera à l'aide de critères comme par exemple celui de l'erreur quadratique entre les observations et les estimations aux points  $x_i$ ,  $i = 1, \dots, n$ , (voir le §5.4). L'estimation de  $f(\cdot)$  qui est la meilleure, au sens de ces critères, sera retenue.

Lorsqu'une *spécification* de points dans  $\mathcal{X}$  est possible (au sens de l'ajout contrôlé de nouvelles données (point d'observation, observation)), compte-tenu de la remarque 2, nous choisirons des points additionnels permettant de réduire la discrépance. Ces points *spécifiés* pourront, par exemple, être choisis parmi ceux d'une suite à discrépance faible.

*Remarques.* — Le cadre méthodologique défini ci-dessus consiste à illustrer l'utilité des inégalités (14) et (38). C'est pourquoi, la sélection des points d'observation  $\mathbf{x}(n) = \{x_1, \dots, x_n\}$  est réalisée de façon à réduire leur discrépance améliorant ainsi, par définition de ce critère, «l'uniformité de leur répartition ». La sélection est donc indépendante des observations lorsque celles-ci sont disponibles (voir le cas 4).

Une perspective envisageable consisterait à utiliser la discrépance dans l'espace des observations, par le calcul de  $D^{L^2}(\mathbf{y}(n))$ ,  $\mathbf{y}(n) = \{y_1, \dots, y_n\}$ , et/ou l'espace défini par le produit de l'espace des points d'observation et l'espace des observations, par le calcul de  $D^{L^2}(\mathbf{x}(n), \mathbf{y}(n))$ . On sélectionnerait les points de façon à améliorer leur répartition uniforme dans ces espaces (i.e., ceux diminuant la discrépance). Ceci permettrait peut-être «d'améliorer» également l'estimation de la fonction (en un sens qui serait à définir). Des justifications théoriques à cette méthode sont des travaux de recherche possibles.

## 5. Application

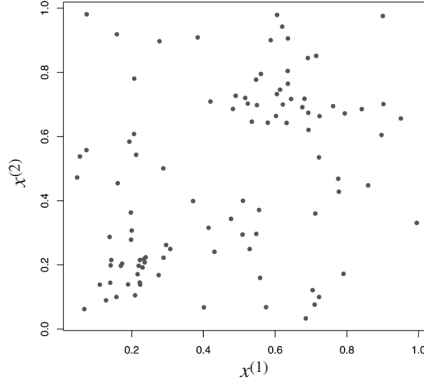
### 5.1. Présentation de l'exemple

Pour l'application, le cas 4 de la méthodologie est abordé ci-dessous. Des observations  $y_i \in \mathbb{R}$  associés à des points  $x_i \in \mathcal{X} = [0, 1]^2$ ,  $i = 1, \dots, 100$  sont disponibles. L'ensemble  $\mathbf{x}(100) = \{x_1, \dots, x_{100}\}$  est représenté dans la Figure 1.

Les observations  $y_1, \dots, y_{100}$  correspondent à la modélisation :

$$y(x_i) = f(x_i) + \varepsilon_i, \quad \text{avec,} \quad (40)$$

- $x_i \in \mathcal{X} = [0, 1]^2$ ,
- $f \in L^2(\mathcal{X})$ , le paramètre fonctionnel à estimer,
- $\varepsilon_i$ ,  $i = 1, \dots, 100$ , une suite de variables aléatoires indépendantes de moyenne nulle de variance  $\sigma^2$ .


 FIG 1. — Ensemble  $\mathbf{x}(100)$  initial.

La fonction  $f(\cdot)$  considérée pour cette application est une somme de produits tensoriels de polynômes de Legendre et s'écrit comme suit,

$$\begin{aligned} \text{pour } x^{(j)} \in [0, 1) : \quad & \phi_0(x^{(j)}) = 1, \quad \phi_1(x^{(j)}) = 12(x^{(j)} - 1/2), \\ \text{pour } x = (x^{(1)}, x^{(2)}) \in \mathcal{X} = [0, 1)^2 \end{aligned}$$

$$\begin{aligned} f(x^{(1)}, x^{(2)}) = & a_{0,0} \phi_0(x^{(1)})\phi_0(x^{(2)}) + a_{0,1} \phi_0(x^{(1)})\phi_1(x^{(2)}) \\ & + a_{1,0} \phi_1(x^{(1)})\phi_0(x^{(2)}) + a_{1,1} \phi_1(x^{(1)})\phi_1(x^{(2)}). \end{aligned}$$

Les coefficients  $a_{0,0}$ ,  $a_{0,1}$ ,  $a_{1,0}$ , et  $a_{1,1}$ , sont estimés par la méthode des moments introduite par la formule (3).

Les variables aléatoires  $\varepsilon_i$ ,  $i = 1, \dots, 100$ , utilisées pour la simulation sont indépendantes et de loi normale centrée  $\mathcal{N}(0, \sigma^2)$ , d'écart type  $\sigma = 0.4$ .

## 5.2. Analyse initiale des points disponibles

Selon la méthodologie définie au paragraphe 4, il est nécessaire de commencer par étudier la «qualité» de l'ensemble  $\mathbf{x}(n) = \{x_1, \dots, x_n\}$ . Pour ce faire la discrédance de type  $L^2$  centrée sera utilisée comme critère de qualité dont le calcul est donné par la formule (voir Hickernell (1998)),

$$\begin{aligned} [\text{DC}^{L^2}(\mathbf{x}(n))]^2 = & \left(\frac{13}{12}\right)^d - \frac{2}{n} \sum_{i=1}^n \prod_{k=1}^d \left(1 + \frac{1}{2}|1 + x_i^{(k)} - 1/2| - \frac{1}{2}|1 + x_i^{(k)} - 1/2|^2\right) \\ & + \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n \prod_{k=1}^d \left(1 + \frac{1}{2}|1 + x_i^{(k)} - 1/2| + \frac{1}{2}|1 + x_j^{(k)} - 1/2| \right. \\ & \quad \left. - \frac{1}{2}|x_i^{(k)} - x_j^{(k)}|\right), \end{aligned}$$

avec ici,  $d = 2$ ,  $n = 100$ , et  $x_i = (x_i^{(1)}, x_i^{(2)}) \in [0, 1)^2$ .

Par définition, la discrédance s'interprète comme une comparaison entre le nombre de points compris dans certains pavés de  $\mathcal{X} = [0, 1]^d$  et le volume de ces pavés. Il s'agit donc d'un critère de répartition uniforme de l'ensemble des points  $\mathbf{x}(n) = \{x_1, \dots, x_n\}$  dans  $\mathcal{X} = [0, 1]^2$ . On parle encore de critère de « remplissage de l'espace » ou « *space filling* ».

Pour que l'ensemble  $\mathbf{x}(n)$  soit jugé de qualité « acceptable », une première approche consisterait à effectuer des *tests statistiques*. L'hypothèse  $H_0$  à tester serait : « les  $x_i$  sont des réalisations de variables aléatoires indépendantes et de même loi uniforme sur  $\mathcal{X} = [0, 1]^d$  », et la statistique utilisée serait :  $DC^{L^2}(\mathbf{x}(n))^2$ . A notre connaissance, la loi de cette statistique n'est pas connue de façon exacte à ce jour. Il n'est donc pas possible d'effectuer de tels tests. Cependant son espérance (Hickernell (1996) et Hickernell (1998)) se calcule et vaut

$$\mathbb{E} \left[ DC^{L^2}(\mathbf{x}(n))^2 \right] = \frac{1}{n} \left[ \left( \frac{13}{12} + \frac{1}{6} \right)^d - \left( \frac{13}{12} \right)^d \right]. \quad (41)$$

Puisque, par définition, plus la discrédance d'un ensemble de points  $\mathbf{x}(n) = \{x_1, \dots, x_n\}$  dans un espace  $\mathcal{X} = [0, 1]^d$  est faible, meilleure est la qualité du recouvrement des points  $\mathbf{x}(n)$  dans  $\mathcal{X}$ , l'espérance (41) constituera notre valeur seuil *supérieure* de référence. Nous jugerons de « qualité acceptable » un ensemble de points  $\mathbf{x}(n)$  dont  $DC^{L^2}(\mathbf{x}(n))^2 < E \left[ DC^{L^2}(\mathbf{x}(n))^2 \right]$ . Notons que, selon nos simulations, lorsque  $\mathbf{x}(n)$  correspond à un ensemble de v.a. i.i.d uniformes,  $\mathbb{E} \left[ DC^{L^2}(\mathbf{x}(n))^2 \right]$  correspond au quantile à 61% de  $DC^{L^2}(\mathbf{x}(n))^2$  :

$$\mathbb{P} \left( DC^{L^2}(\mathbf{x}(n))^2 \leq \mathbb{E} \left[ DC^{L^2}(\mathbf{x}(n))^2 \right] \right) \approx 0.61.$$

La densité de la statistique  $n \times DC^{L^2}(\mathbf{x}(n))^2$  obtenue par simulation est représentée dans la Figure 2 (la multiplication par  $n$  permet de ne pas faire dépendre la loi de la statistique du nombre de points).

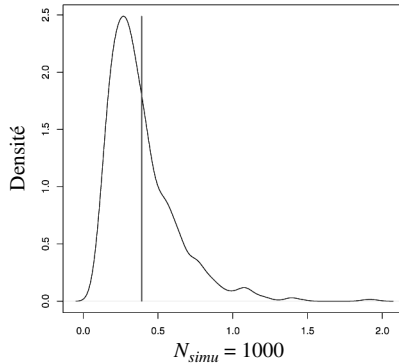


FIG 2. — Densité de  $n \times DC^{L^2}(\mathbf{x}(n))^2$ , la moyenne est représentée par le trait vertical.



Pour l'ensemble initial  $\mathbf{x}(100) = \{x_1, \dots, x_{100}\}$ , nous avons

$$[\text{DC}^{L^2}(\mathbf{x}(n))]^2 = 0.0100 > \mathbb{E} \left[ \text{DC}^{L^2}(\mathbf{x}(n))^2 \right] = 0.0038.$$

L'ensemble  $\mathbf{x}(100) = \{x_1, \dots, x_{100}\}$  n'est donc pas jugé de qualité acceptable. La Figure 11 montre aussi que l'ajustement de l'estimation de  $f(\cdot)$  par la méthode de projection sur une base de fonctions orthonormales n'est pas satisfaisant lorsque les points de  $\mathbf{x}(100) = \{x_1, \dots, x_{100}\}$  sont pris en compte.

### 5.3. Sélection d'un sous-ensemble de points

Selon le cas 4 de la méthodologie introduite au paragraphe 4, nous proposons de sélectionner un sous-ensemble  $\mathbf{x}_1(n_1) \subset \mathbf{x}(n) = \{x_1, \dots, x_n\} \in \mathcal{X}^n$  de façon à réduire la quantité

$$Q(\mathbf{x}(n)) = \frac{\text{DC}^{L^2}(\mathbf{x}(n))}{n}. \quad (42)$$

Ici,  $\text{DC}^{L^2}(\mathbf{x}(n))$  désigne la discrédance de type  $L^2$  centrée. Pour effectuer cette sélection, plusieurs méthodes sont possibles et vont être illustrées.

#### 5.3.1. Méthode 1

La première méthode consiste à trouver un sous-ensemble  $\mathbf{x}_1(n_1)$  de  $\mathbf{x}(100)$  dont  $Q(\mathbf{x}(n_1))$  (cf. (42)) est minimale. L'algorithme  $A_1$  utilisé est le suivant :

Algorithme  $A_1$

Étape 1 : calcul de  $D_{\text{iter}_0} = Q(\mathbf{x}(n))$ ;

Étape 2 : Pour  $i = 1, \dots, n$ ,

- Calcul de  $D_{\text{iter}_1, i} = Q(\mathbf{x}_{-i}(n-1))$   
où  $\mathbf{x}_{-i}(n-1) = \{x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_n\}$ ,
- Calcul de  $\text{diff}_i = D_{\text{iter}_1, i} - D_{\text{iter}_0}$ ;

Étape 3 : Si  $\text{diff}_{i^*} = \{\min_{i=1, \dots, n} \text{diff}_i\} < 0$ ,

- sélection de l'ensemble  $\mathbf{x}_{-i^*}(n-1)$ ,
- itération (retour à l'Étape 1) en posant :  
 $\mathbf{x}(n) \leftarrow \mathbf{x}_{-i^*}(n-1)$ ,  $D_{\text{iter}_0} \leftarrow D_{\text{iter}_1}$ ;

Étape 4 : Arrêt lorsque  $D_{\text{iter}_1, i^*} = \min_{i=1, \dots, n} D_{\text{iter}_1, i} > 0$ .

Cet algorithme ne permet pas forcément de trouver le sous-ensemble optimal, puisqu'on itère en considérant uniquement la différence  $\text{diff}_i = D_{\text{iter}_1, i} - D_{\text{iter}_0}$ . Cependant, ce sous-ensemble peut être qualifié «  $Q(\mathbf{x}(n))$  - irréductible », puisqu'il y a arrêt de l'algorithme lorsqu'aucune suppression de points ne permet de réduire cette quantité. Lors de l'itération, l'ensemble de points

$\mathbf{x}_r(n_r) = \{x_i : \text{diff}_i < 0\} \subset \mathbf{x}(n)$  constitue un ensemble de points qui ne contribue pas à une répartition uniforme de l'ensemble  $\mathbf{x}(n)$ . En effet, la suppression d'un de ces points permet de diminuer le critère  $Q$ , et donc la discrépance, critère de répartition uniforme (voir §2 et les équations (39)). L'ensemble  $\mathbf{x}_r(n_r) = \{x_i : \text{diff}_i < 0\} \subset \mathbf{x}(n)$  peut donc être qualifié d'ensemble de points redondants. L'ensemble  $\mathbf{x}_r(n_r) = \{x_i : \text{diff}_i < 0\}$  initial est représenté par le graphique 3 (points encerclés). Ce n'est pas pour autant un ensemble redondant, car seule la suppression d'un unique point de l'ensemble permet de réduire la discrépance. Il n'y a aucune garantie pour que la suppression de l'ensemble  $\mathbf{x}_r(n_r)$  permette de réduire  $Q$ . Pour ce faire, il faudrait vérifier que :

$$\text{diff}_{i_1, \dots, i_{n_r}} = Q(\mathbf{x}(n)) - Q(\mathbf{x}(n) \setminus \mathbf{x}_r(n_r)) < 0.$$

L'évolution de la quantité  $Q$  en fonction du nombre de points supprimés est représentée Figure 5. La courbe en vert représente  $\sqrt{\mathbb{E} [\text{DC}^{L^2}(\mathbf{x}(n))^2]} / n$  qui constitue une valeur seuil de référence (voir §5.2). L'ensemble de points retenu sera celui dont la quantité  $Q$  est la plus faible. Il comporte 54 points. On remarque toutefois, qu'à partir de 80 points il est possible de sélectionner un ensemble de points de qualité « acceptable ».

L'algorithme  $A_1$  est réalisable car le nombre de points ( $n = 100$ ) et la dimension de l'espace ( $d = 2$ ) considérés pour l'application sont relativement faibles. Le temps de calcul pour obtenir ces résultats est pratiquement négligeable. Lorsque le nombre de points et la dimension sont plus importants, l'exécution de cet algorithme devient plus couteuse en temps de calcul. Il est donc nécessaire de le modifier quelque peu. Par exemple, à l'étape 2, une méthode consiste à remplacer  $D_{\text{iter}_1, i}$  par  $D_{\text{iter}_1, \hat{i}_k} = Q(\mathbf{x}_{-\hat{i}_k}(n - k))$  où  $\mathbf{x}_{-\hat{i}_k}(n - k) = \{x_1, \dots, x_n\} \setminus \{x_{i_1}, \dots, x_{i_k}\}$ , pour un ensemble d'indices  $\hat{i}_k = \{i_1, \dots, i_k\} \subset \{1, \dots, n\}$ , puis à arrêter la boucle dès que  $\text{diff}_{\hat{i}_k} = D_{\text{iter}_1, \hat{i}_k} - D_{\text{iter}_0} < 0$ .

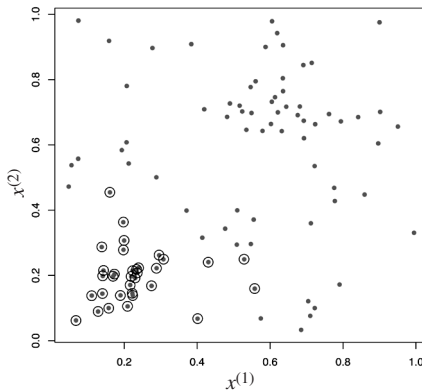
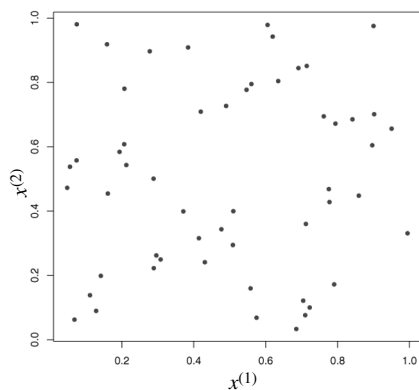
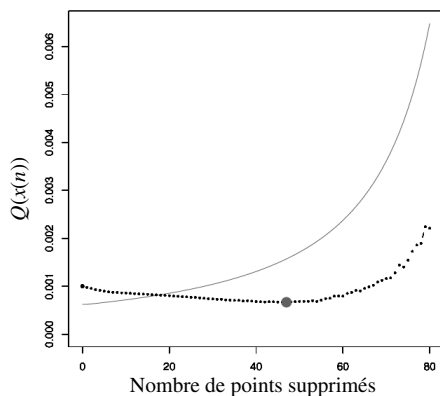


FIG 3. — Groupe de points « redondants » (encerclés) à l'étape initiale.

FIG 4. — Points sélectionnés par  $A_1$ .FIG 5. — Évolution de  $Q$  en fonction du nombre de points supprimés par  $A_1$ .

### 5.3.2. Méthode 2

L'objectif de la seconde méthode est d'extraire un sous-ensemble de points régulièrement répartis dans  $\mathcal{X} = [0, 1]^2$ . L'algorithme  $A_2$  est le suivant.

Le fait de minimiser le critère  $Q$  (cf. (42)) dans cet algorithme permet aussi de diminuer la discrédance, critère de répartition uniforme (voir §2 et les équations (39)). Rappelons que la discrédance s'interprète géométriquement comme une comparaison entre le nombre de points de  $\mathbf{x}(n)$  compris dans certains pavés de  $\mathcal{X} = [0, 1]^2$  et le volume de ces pavés. Ainsi, nous pouvons raisonnablement penser qu'un ensemble de points régulièrement espacés permettra d'avoir une discrédance faible. C'est entre autres le cas des réseaux ou «*lattice*» (nous renvoyons par exemple à Lemieux et L'Ecuyer (2001)). L'objectif de l'algorithme  $A_2$  est donc de sélectionner des points régulièrement

espacés, tous distants d'au moins  $\varepsilon > 0$ , et qui permettent de recouvrir au mieux l'espace  $\mathcal{X} = [0, 1]^2$ . Cet algorithme peut aussi être modifié en remplaçant à l'étape 3 :  $x_{i'} \in \mathbf{x}_{\text{iter}_2}(n - n_r - 1)$ , le point le plus proche de  $\mathbf{x}_{\text{iter}_1}(1)$ , par  $x_{i'} \in \mathbf{x}_{\text{iter}_2}(n - n_r - 1)$  le point le plus éloigné de  $\mathbf{x}_{\text{iter}_1}(1)$ . Ce dernier réalise le *maximum* des distances euclidiennes entre les points de  $\mathbf{x}_2(n - 1 - n_r)$  et  $\mathbf{x}_{\text{iter}}(1)$ . L'ensemble de points alors retenu à l'aide de l'algorithme  $A_2$  modifié comporte des caractéristiques similaires à celui retenu par l'algorithme  $A_2$  initial.

### Algorithme $A_2$

Étape 1 : Soit  $\text{dist} = \{\varepsilon_1, \dots, \varepsilon_{\text{end}}\}$  un ensemble de valeurs croissantes dans  $(0, 1)$ .

Sélection du point  $x_i^* \in \mathbf{x}(n)$  le plus proche (au sens de la norme euclidienne) du « centre » du domaine (i.e. du point  $(0.5, 0.5)$ ).

- $\mathbf{x}_{\text{iter}_1}(1) = \{x_i^*\}$ ,
- $\mathbf{x}_{\text{iter}_2}(n - 1) = \mathbf{x}(n) \setminus \{x_i^*\}$ .

Étape 2 : Soient  $\mathbf{x}_r(n_r) = \{x_{i_1}, \dots, x_{i_{n_r}}\} \subset \mathbf{x}_{\text{iter}_2}(n - 1)$  les points inclus dans la boule de centre  $x_i^*$  et de rayon  $\varepsilon_1$  :  $\mathbf{x}_r(n_r) = \mathbf{x}_{\text{iter}_2}(n - 1) \cap B(x_i^*, \varepsilon_1)$ .

- $\mathbf{x}_{\text{iter}_2}(n - n_r - 1) = \mathbf{x}_{\text{iter}_2}(n - 1) \setminus \mathbf{x}_r(n_r)$ .

Étape 3 : Soit  $x_{i'} \in \mathbf{x}_{\text{iter}_2}(n - n_r - 1)$  le point le plus proche de  $\mathbf{x}_{\text{iter}_1}(1)$  (réalisant le minimum des distances entre les points de  $\mathbf{x}_2(n - 1 - n_r)$  et  $\mathbf{x}_{\text{iter}}(1)$ ).

- $\mathbf{x}_{\text{iter}_1}(2) = \mathbf{x}_{\text{iter}}(1) \cup \{x_{i'}\}$
- $\mathbf{x}_{\text{iter}_2}(n - n_r - 2) = \mathbf{x}_{\text{iter}_2}(n - n_r - 1) \setminus \{x_{i'}\}$ .

Étape 4 : Soient  $\mathbf{x}_{r_2}(n_{r_2}) \in \mathbf{x}_{\text{iter}_2}(n - n_r - 2)$  les points inclus dans la boule de centre  $x_{i'}$  et de rayon  $\varepsilon_1$  :  $\mathbf{x}_{r_2}(n_{r_2}) = \mathbf{x}_{\text{iter}_2}(n - n_r - 2) \cap B(x_{i'}, \varepsilon_1)$ .

- $\mathbf{x}_{\text{iter}_2}(n - n_r - 2) = \mathbf{x}_{\text{iter}_2}(n - n_r - n_{r_2} - 2) \setminus \mathbf{x}_{r_2}(n_{r_2})$ .

Étape 5 : Itération des Étapes 3 et 4 tant que  $\mathbf{x}_{\text{iter}_2} \neq \emptyset$ .

On note  $\mathbf{x}_{\varepsilon_1}(n_{\varepsilon_1})$  l'ensemble  $\mathbf{x}_{\text{iter}_1}(n_1)$  obtenu.

Étape 6 : Calcul de  $Q(\mathbf{x}_{\varepsilon_1}(n_{\varepsilon_1}))$ . Retour à l'Étape 1 avec  $\varepsilon_2 \in \text{dist} = \{\varepsilon_1, \dots, \varepsilon_{\text{end}}\}$ .

Étape 7 : Sélection de  $\mathbf{x}_1(n_1) = \mathbf{x}_\varepsilon(n_\varepsilon) = \arg_{\varepsilon_i \in \text{dist}} \{\min Q(\mathbf{x}_{\varepsilon_i}(n_{\varepsilon_i}))\}$ .

Les Figures 7 et 8 représentent, l'évolution de la discrédance en fonction du nombre de points supprimés, et de la distance  $\varepsilon$  fixée. L'ensemble de points retenu comporte 48 points et est illustré Figure 6. Nous remarquons qu'à partir de 75 points il est possible d'obtenir un ensemble de points de qualité acceptable selon nos critères (voir §5.2).

En général, cet algorithme permet de trouver un sous-ensemble de points de qualité acceptable. Le critère  $Q$  diminue, cependant, moins rapidement et moins fortement que lors de l'application de l'algorithme  $A_1$  (voir les Figures 5 et 7).

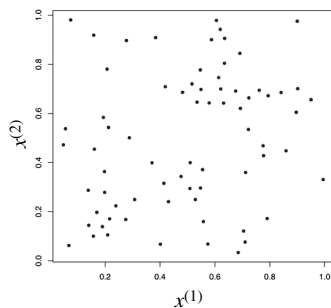


FIG 6. — Points sélectionnés par  $A_2$ .

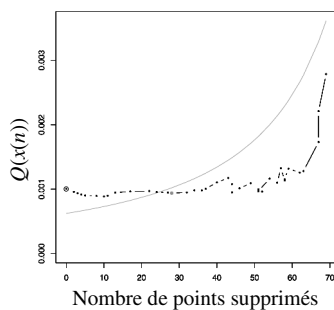


FIG 7. — Évolution de  $Q$  en fonction du nombre de points supprimés par  $A_2$ .

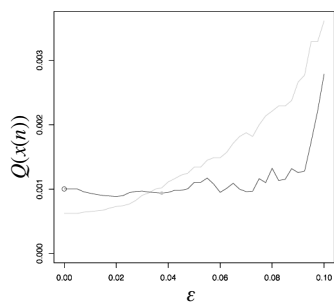


FIG 8. — Évolution de  $Q$  en fonction de la distance  $\varepsilon$  par  $A_2$ .

### 5.3.3. Méthode 3

La troisième méthode consiste simplement à sélectionner des points proches de ceux d'une *suite à discrédance faible*. Ces suites «déterministes» sont essentiellement utilisées pour les méthodes de quasi-Monte Carlo dont l'objectif est de réduire au mieux la discrédance à l'origine (voir le §2). Pour l'étude de ces suites, nous renvoyons, par exemple, à Niederreiter (1998).

Dans l'algorithme  $A_3$  présenté ci-dessous, le critère  $Q$  est défini en (42).

#### Algorithme $A_3$

Étape 1 : Construction d'une *suite à discrédance faible* dans  $\mathcal{X}$  comportant  $n_{sdf}$  points,  $\mathbf{x}_{sdf}(n_{sdf}) = \{x_{sdf,1}, \dots, x_{sdf,n_{sdf}}\}$ .

Étape 2 : Pour  $i = 1, \dots, n_{sdf}$ ,

- sélection de  $x_{i_1} \in \mathbf{x}(n)$  le plus proche du point  $x_{sdf,i} \in \mathbf{x}_{sdf}(n_{sdf})$ ,
- $\mathbf{x}(n-1) = \mathbf{x}(n) \setminus x_{i_1}$ ,
- $\mathbf{x}_{sdf}(n_{sdf}-1) = \mathbf{x}_{sdf}(n_{sdf}) \setminus x_{sdf,i}$ .

Soit  $\mathbf{x}(n_{sdf}) \subset \mathbf{x}(n)$  la suite obtenue.

Calcul de  $Q(\mathbf{x}(n_{sdf}))$ .

Étape 3 : Construction d'une suite à discrédance faible comportant  $n_{sdf'}$  points,  $\mathbf{x}_{sdf'}(n_{sdf'}) = \{x_{sdf',1}, \dots, x_{sdf',n_{sdf'}}\}$  avec  $n_{sdf'} < n_{sdf}$ .

Application de l'Étape 2.

L'avantage de cette méthode simple est qu'elle permet d'obtenir immédiatement un ensemble ayant un nombre de points souhaité sans réaliser d'itérations (de calculs répétés de la discrédance). Elle peut donc se révéler aussi très efficace lors du cas 2 de la méthodologie du §4.

L'évolution de la discrédance en fonction du nombre de points supprimés est présentée dans la Figure 10. A partir de 80 points, il est possible d'obtenir un ensemble de qualité acceptable selon nos critères (voir le §5.2). L'ensemble de points retenu par cet algorithme comporte 67 points. Il est illustré dans la Figure 9. Entre 67 points et 45 points, la valeur de  $Q$  obtenue évolue relativement peu.

## 5.4. Estimation et Validation

Comparons à présent les différentes estimations de  $f(\cdot)$ , désignées par  $\hat{f}_{ini}(\cdot)$ ,  $\hat{f}_{A_1}(\cdot)$ ,  $\hat{f}_{A_2}(\cdot)$ ,  $\hat{f}_{A_3}(\cdot)$ , obtenues, respectivement, avec l'ensemble  $\mathbf{x}(n)$  de points initiaux, l'ensemble  $\mathbf{x}_{A_1}(n_1)$  de points sélectionnés par l'algorithme  $A_1$ , l'ensemble  $\mathbf{x}_{A_2}(n_2)$  de points sélectionnés par l'algorithme  $A_2$ , l'ensemble  $\mathbf{x}_{A_3}(n_3)$  de points sélectionnés par l'algorithme  $A_3$ .

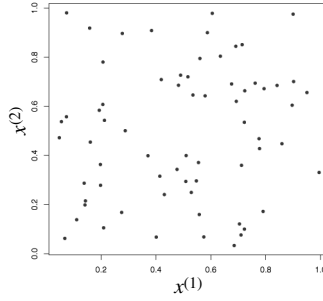


FIG 9. — Points sélectionnés par  $A_3$ .

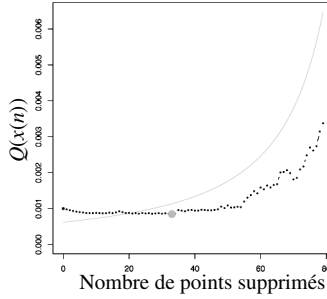


FIG 10. — Évolution de  $Q$  en fonction du nombre de points supprimés par  $A_3$ .

Pour la validation des estimations de  $f(\cdot)$ , nous avons introduit un nouvel ensemble  $\mathbf{x}_{val}(100)$  de 100 points correspondant à des réalisations de variables aléatoires indépendantes uniformes dans  $[0, 1]^2$ . Pour  $i = 1, \dots, 100$ , des observations  $y_{val_i} = f(x_i) + \varepsilon_i$ , où  $\varepsilon_i \sim \mathcal{N}(0, 0.4)$ , ont été simulées aux points  $x_{val_i} \in \mathbf{x}_{val}(100)$ . Les Figures 11, 12, 13, 14, représentent, respectivement, les estimations  $\hat{f}_{ini}(x_{val_i})$ ,  $\hat{f}_{A_1}(x_{val_i})$ ,  $\hat{f}_{A_2}(x_{val_i})$ ,  $\hat{f}_{A_3}(x_{val_i})$ , de  $f(x_{val_i})$ , rapportés aux observations «de validation»,  $y_{val_i}$ ,  $i = 1, \dots, 100$ . Le Tableau 15 récapitule l'ensemble des résultats obtenus lors de l'étape de sélection de points, ainsi que l'erreur quadratique estimée par la formule

$$EQ_{A_j} = \frac{1}{100} \sum_{x_i \in \mathbf{x}_{val}(100)} \left( \hat{f}_{A_j}(x_i) - y_i \right)^2 \quad \text{pour } A_j = A_1, A_2, A_3.$$

Le Tableau 15 montre que l'algorithme  $A_1$  donne de meilleurs résultats que les algorithmes  $A_2$  et  $A_3$ . Ceci n'est pas surprenant car l'objectif de  $A_1$  est la minimisation directe du critère  $Q$ .

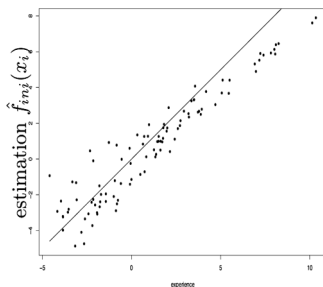


FIG 11. — Validation de l'estimation  $\hat{f}_{ini}(\cdot)$ .

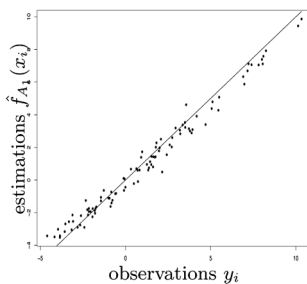


FIG 12. — Validation de l'estimation  $\hat{f}_{A_1}(\cdot)$ .

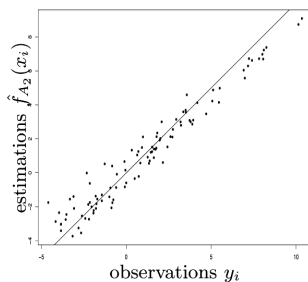


FIG 13. — Validation de l'estimation  $\hat{f}_{A_2}(\cdot)$ .

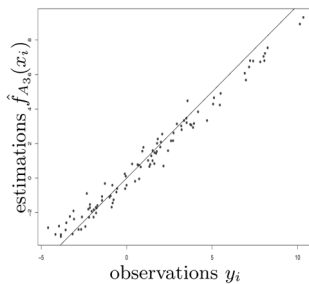


FIG 14. — Validation de l'estimation  $\hat{f}_{A_3}(\cdot)$ .



Ensemble de points	Nb. points	$DC^{L^2}$	$\mathbb{E}(DC^{L^2})$	$Q$	EQ
$\mathbf{x}_{A_1}(n_1)$	54	1.3E-3	7.2E-3	6.6E-4	0.22
$\mathbf{x}_{A_2}(n_2)$	72	4.6E-3	5.4E-3	9.4E-4	0.66
$\mathbf{x}_{A_3}(n_3)$	67	3.0E-3	5.8E-3	8.1E-4	0.34
$\mathbf{x}(n)$	100	10E-2	3.8E-3	10E-4	1.37

TABLEAU 15. — Tableau comparatif des résultats des différents algorithmes et des ajustements associés

Comme l'objectif principal de l'algorithme  $A_2$  n'est pas la minimisation du critère  $Q$  mais la sélection d'un ensemble de points régulièrement répartis dans  $\mathcal{X}$ , celui-ci donne de moins bons résultats. De plus, le nombre de points sélectionnés par l'algorithme  $A_2$  est plus important que ceux des algorithmes  $A_1$  et  $A_3$ . Ainsi, l'ensemble  $\mathbf{x}_{A_2}(n_2)$  conservera un nombre relativement important de points de l'ensemble initial  $\mathbf{x}(n)$ . Comme le caractère de répartition uniforme de l'ensemble initial n'est pas jugé « acceptable » selon nos critères, le fait de sélectionner, parmi cet ensemble, une partie de cardinalité trop élevée risque de conserver ce manque d'uniformité. Ainsi  $\mathbf{x}_{A_2}(n_2)$ , risquera lui aussi de ne pas avoir une bonne « répartition ». Ceci est notamment confirmé par un critère sévère d'uniformité tel que celui de la discrédance  $L^2$  centrée, voir le Tableau 15.

L'algorithme  $A_3$  permet, en général, d'extraire un ensemble de points « proche » d'une suite à *discrédance faible* et donc de bénéficier parfois de caractéristiques semblables (recouvrement uniforme de l'espace). Il est également moins coûteux en temps de calcul, ce qui le rend particulièrement intéressant lorsque la dimension de l'espace  $\mathcal{X}$  est élevée.

La sélection d'un ensemble de points par minimisation du critère  $Q = DC^{L^2}(\mathbf{x}(n))/n$  prend en compte l'inégalité (14). Comme précisé au §4, la division par  $n$  contraint la sélection à conserver un nombre de points relativement important. Bien entendu, il est aussi possible d'appliquer cette méthodologie en cherchant simplement à minimiser la discrédance centrée  $DC^{L^2}$ . Le nombre de points sélectionné sera alors inférieur à celui où traité par le critère  $Q$ . Les résultats obtenus seront alors comparables.

## 6. Conclusion, Perspectives

En nous appuyant sur les travaux de Hickernell (1998) et Rafajlowicz et Schwabe (2006) nous avons précisé les liens entre les critères d'IMSE et de MSE et la notion de discrédance généralisée, dans le cadre de l'estimation d'un paramètre fonctionnel basée sur sa décomposition à partir d'une somme de fonctions orthonormales. Les critères d'IMSE et de MSE peuvent tous deux être majorés par un terme faisant intervenir la discrédance généralisée. Ainsi,

une « discrédance » faible de l'ensemble de points  $\mathbf{x}(n) = \{x_1, \dots, x_n\}$  en lesquels sont observés  $\{y_1, \dots, y_n\}$  permet d'obtenir une estimation *robuste*, au sens de la MISE, et de bonne qualité au sens de l'IMSE. C'est l'objectif poursuivi par la méthodologie proposée, et illustrée par une application.

La réduction d'un critère faisant uniquement intervenir les points en lesquels seront réalisées des observations peut également s'inscrire dans le contexte des plans d'expérience. L'utilisation de tels critères est classique en régression linéaire (voir par exemple l'ouvrage de Dreesbeke *et al.* (1997)). Ceux-ci sont exactement proportionnels à des critères de qualité de la régression (considération d'égalités et non de majorations). Les critères de plan d'expériences sont cependant exploitables dans des contextes de régression bien définis et ne s'appliquent pas à la méthode des fonctions orthogonales utilisée ici.

Plus généralement, la discrédance étant un critère de répartition uniforme, la méthodologie que nous avons mise en place peut être considérée comme une méthode, dite de « space-filling design », où l'objectif est de vérifier que des points recouvrent de façon uniforme l'espace dans lequel ils sont définis. Cette méthode peut donc être utilisée dans un contexte plus large.

Le lien entre la discrédance et la méthode des fonctions orthogonales a été établi à l'aide de l'inégalité de Koksma-Hlwwaka généralisée. Lorsqu'une méthode d'estimation fera intervenir des moyennes, approximations d'intégrale, cette inégalité montre que la diminution de la discrédance d'un ensemble de points permettra de réduire l'erreur de l'estimation. Cette propriété pourrait aussi être employée dans le contexte de l'apprentissage statistique où l'on considère une fonction de risque empirique, approximation d'une intégrale (ceci fait l'objet de recherches récentes, voir J. Marry (2005) et Cervellera et Muselli (2004)). Lorsqu'il est possible de spécifier des données d'entrées (que nous appelons ici points d'observation) pour effectuer la procédure d'apprentissage, Cervellera et Muselli (2004) recommandent notamment l'utilisation de suites à discrédance faible. Il semblerait également pertinent d'utiliser des algorithmes d'extraction et/ou de spécification de points comme ceux présentés au §5.3 dans la procédure d'apprentissage.

D'origine récente (Hickernell (1998)), l'inégalité de Koksma-Hlwwaka généralisée permet de définir la discrédance à partir d'un noyau auto-reproduisant convenable d'un espace de Hilbert. Lorsqu'une modélisation par un processus aléatoire est utilisée, cela revient aussi à considérer des espaces de Hilbert où le noyau auto-reproduisant correspond à la fonction de covariance du processus. Appliquer la notion de discrédance en adaptant la définition à la fonction de covariance comme mentionné par Hickernell (1999) constitue également une perspective de recherche intéressante.

## 7. Remerciements

Nous remercions le professeur P. Deheuvels (Université Paris VI), N. Devictor et B. Iooss du CEA de Cadarache, le professeur R. Phan-Tan-Luu (Université Aix-Marseille III), pour leur aide précieuse et leurs relectures avisées.

## Bibliographie

- J. B. AUBIN. *Estimation fonctionnelle par projection adaptative et applications*. Thèse de doctorat, Université Pierre et Marie Curie, Paris VI, 2005.
- D. BOSQ. Sur l'estimation d'une densité multivariée par une série de fonctions orthogonales. *Comptes Rendus de l'Académie des Sciences de Paris*, 268 : 555-557, 1969.
- D. BOSQ et J. BLUEZ. Etude d'une classe d'estimateurs non-paramétriques de la densité. *Annales de l'institut Henri Poincaré*, 14 : 479-498, 1978.
- D. BOSQ et J. LECOUTRE. *Théorie de l'estimation fonctionnelle*. Economica, 1987.
- M. CENCOV. Evaluation of an unknown distribution density from observations. *Soviet Mathematics*, 3 : 1559-1562, 1962.
- C. CERVELLERA et M. MUSELLI. Deterministic design for neural network learning : An approach based on discrepancy. *IEEE Transactions on Neural Networks*, 15 : 533-544, 2004.
- L. DEVROYE et L. GYORFI. *Nonparametric Density Estimation The  $L_1$  view*. John Wiley & Sons, Inc., 1985.
- J. DROESBEKE, J. FINE et G. SAPORTA. *Plans d'expériences, Applications à l'entreprise*. Technip, 1997.
- A. DVORETZKY, J. KIEFER et J. WOLFOWITZ. Asymptotic minimax character of the sample distribution function and of a classical multinomial estimator. *Annals of Mathematical Statistics*, 33 : 642-669, 1956.
- V. FEUILLARD, N. DEVICTOR et R. PHAN-TAN-LUU. Liens entre discrédance et estimation non-paramétrique, méthodologie de sélection de points selon les données disponibles. 38<sup>ème</sup> journées de Statistiques, 2006.
- F. HICKERNELL. The mean square discrepancy of randomized nets. *Modeling and Computer Simulation*, 6(4) : 274-296, 1996.
- F. HICKERNELL. A generalized discrepancy and quadrature error bound. *Mathematics of computation*, 67(221) : 299-322, 1998.
- F. HICKERNELL. Goodness-of-fit statistics, discrepancies and robust designs. *Statistics and Probability Letters*, 44 : 73-78, 1999.
- E. HLWAKA. Funktionen von beschränkter Variation in der Theorie der Gleichverteilung. *Annali di Matematica Pura ed Applicata*, 54 : 325-333, 1961.
- W. HÄRDLE. *Applied nonparametric regression*. Cambridge University Press, 1989.
- J. MARRY. *Étude de l'Apprentissage Actif : Application à la conduite d'expérience*. Thèse de doctorat, Université Paris Sud, 2005.
- C. LEMIEUX et P. L'ÉCUYER. On selection criteria for lattice rules and other quasi-monte carlo point sets. *Mathematics and Computers in Simulation*, 55 : 139-148, 2001.
- P. MASSART. The tight constant in the dvoretzky-kiefer-wolfowitz inequality. *Annals of Probability*, 18(3) : 1269-1283, 1990.
- H. NIEDERREITER. *Monte Carlo and Quasi-Monte Carlo Methods*. Springer, 1998.
- B. L. S. PRAKASA RAO. *Nonparametric functional estimation*. Probability and Mathematical Statistics. Academic Press Inc.[Harcourt Brace Jovanovich Publishers], 1983.
- E. RAFAJLOWICZ et R. SCHWABE. Halton and Hammersley sequences in multivariate nonparametric regression. *Statistics and probability letters*, 76(8) : 803-812, 2006.

L'EXPLORATION STATISTIQUE DU BIAIS DE PUBLICATION

- F. RIESZ et B. S. NAGY. *Functional Analysis*. Ungar Publishing Co., 1955.
- G. SANSONE. *Orthogonal Functions*. Krieger, 1977.
- C. STONE. Optimal global rate of convergence for nonparametric regression. *Annals of Statistics*, 10 : 1040-1053, 1982.
- E. THIÉMARD. *Sur le calcul et la majoration de la discrédance à l'origine*. Thèse de doctorat, Ecole Polytechnique Fédérale de Lausanne, 2000.