Noyaux et convolution

Table des matières

Le document ci-dessous présente la notion de noyaux utilisée en estimation non-paramétrique, ainsi que leurs propriétés d'approximation. La plupart des résultats peuvent être trouvés dans les livres de Tsybakov (2009) et Comte (2015).

1 Définitions et construction

1.1 Définition et exemples

Définition 1 *Un noyau est une fontion $K : \mathbb{R}^d \to \mathbb{R}$, intégrable sur $\mathbb{R}^d$, d'intégrale égale à 1.*

Remarque. Si $K_1, \dots, K_d$ sont des noyaux sur $\mathbb{R}$, alors $K : x = {}^t(x_1, \dots, x_d) \mapsto K_1(x_1) \dots K_d(x_d)$ est un noyau sur $\mathbb{R}^d$. Dans la suite de ce document, on considèrera donc des noyaux sur $\mathbb{R}$.

Exemples de noyaux sur $\mathbb{R}$ (voir Figure 1).

(a) Noyau rectangulaire : $x \mapsto (1/2)\mathbf{1}_{|x|\leq 1}$.
(b) Noyau triangulaire : $x \mapsto (1-|x|)\mathbf{1}_{|x|\leq 1}$.
(c) Noyau d'Epanechnikov (parabolique) : $x \mapsto (3/4)(1-x^2)\mathbf{1}_{|x|\leq 1}$.
(d) Noyau 'biweight' : $x \mapsto (15/16)(1-x^2)^2\mathbf{1}_{|x|\leq 1}$.
(e) Noyau gaussien : $x \mapsto (1/\sqrt{2\pi})\exp(-x^2/2)$.
(f) Noyau de Silverman : $x \mapsto (1/2)\exp(-|x|/\sqrt{2})\sin(|x|/\sqrt{2}+\pi/4)$.

Remarque. Tous ces noyaux sont positifs (et pairs), à l'exception du noyau de Silverman.

1.2 Noyaux d'ordre donné

1.2.1 Définition

Définition 2 *Soit l un entier naturel non nul. Un noyau K sur $\mathbb{R}$ est dit d'ordre l, si, pour $j = 0, \dots, l$, les fonctions $x \mapsto x^j K(x)$ sont intégrables, et*

$$\int_{\mathbb{R}} x^j K(x)dx = 0.$$

Exemple. Le noyau gaussien est d'ordre 1.

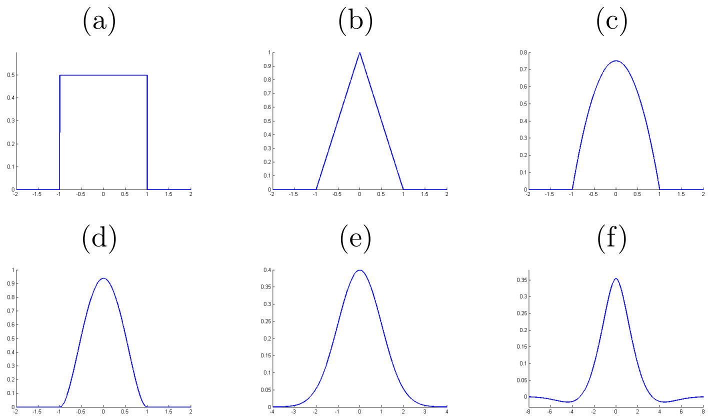


Figure 1 – Principaux noyaux utiles en statistique.

1.2.2 Construction

On peut construire des noyaux d'ordre l donné, au moins à partir de deux méthodes.

Une première construction est par exemple celle proposée par Kerkyacharian *et al.* (2001).

Proposition 3 *Soit $g : \mathbb{R} \to \mathbb{R}$ une fonction intégrable, telle que $\int_{\mathbb{R}} g(x)dx = 1$. Soit $l \in \mathbb{N}\backslash\{0\}$. On définit,*

$$\begin{array}{rccl} K_l : & \mathbb{R} & \longrightarrow & \mathbb{R} \\ & u & \mapsto & K_l(u) = \displaystyle\sum_{k=1}^{l} \binom{l}{k} (-1)^{k+1} \frac{1}{k} g\left(\frac{u}{k}\right). \end{array}$$

Alors,

1. *la fonction K_l est un noyau et si le support de la fonction g est un segment $[a;b]$, celui de K_l est $[a, lb]$;*
2. *si de plus, pour $k \in \{1, \dots, l\}$ les fonctions $u \mapsto u^k g(u)$ sont intégrables, alors K_l est un noyau d'ordre $l-1$.*

Exercice. Le but est de prouver la Proposition 3.

1. Prouver le point (1) de la proposition.
2. Justifier que pour des entiers j, k, l tels que $1 \le j \le k \le l$, on a
$$\binom{l}{k}\binom{k}{j} = \binom{l}{j}\binom{l-j}{k-j}.$$
En particulier, si $j = 1$, on retrouve :
$$k\binom{l}{k} = l\binom{l-1}{k-1}.$$
3. Montrer que pour tous les entiers j, l tels que $1 \le j < l$, on a
$$\sum_{k=1}^{l} (-1)^k \binom{l}{k}\binom{k}{j} = 0.$$

4. En déduire que pour tous les entiers j, l tels que $1 \leq j < l$

$$\sum_{k=1}^{l} \binom{l}{k} (-1)^k Q_j(k) = 0 \text{ avec } Q_j(X) = X(X-1)(X-2)\cdots(X-j+1). \tag{1}$$

5. Prouver par "récurrence" (finie) sur j que

$$\forall j \in \{1, \ldots, l-1\}, \sum_{k=1}^{l} \binom{l}{k} (-1)^k k^j = 0.$$

6. En déduire que $\int_{\mathbb{R}} x^j K_l(x) dx = 0$ pour $j \in \{1, \ldots, l-1\}$ (ce qui termine la preuve de la Proposition 3).

La seconde construction repose sur la base des polynômes de Legendre.

Définition 4 *On appelle polynômes de Legendre les fonctions $(Q_n)_{n \in \mathbb{N}}$ suivantes définies pour $x \in [-1; 1]$:*

$$Q_0(x) = \frac{1}{\sqrt{2}}, \;\; Q_n(x) = \sqrt{\frac{2n+1}{2}} \frac{1}{2^n (n!)} \frac{d^n}{dx^n} \left[(x^2 - 1)^n \right], \;\; n \geq 1.$$

Proposition 5 *La famille $(Q_n)_{n \geq 1}$ des polynômes de Legendre forme une base hilbertienne de $L^2([-1, 1])$, échelonnée en degré (Q_n est de degré n pour tout $n \in \mathbb{N}$).*

La preuve de ce résultat est l'objet de l'Exercice 6.2.8, Chap. 6 du livre de Lacombe et Massat (1999). La construction suivante est suggérée par Tsybakov (2009).

Proposition 6 *Soit $l \in \mathbb{N} \backslash \{0\}$. On définit,*

$$\begin{array}{rccl} K_l : & \mathbb{R} & \longrightarrow & \mathbb{R} \\ & u & \mapsto & K_l(u) = \displaystyle\sum_{m=0}^{l} Q_m(0) Q_m(u) \mathbf{1}_{|u| \leq 1}. \end{array}$$

Alors, la fonction K_l est un noyau d'ordre l.

Exercice. Prouver la Proposition 6. *Indication :* pour calculer $\int_{\mathbb{R}} u^j K_l(u) du$, on pourra décomposer le monôme $P_j(u) = u^j$ dans la base de Legendre, et utiliser le fait que celle-ci est orthonormée.

2 Propriétés d'approximation

Rappel. Soient f et g deux fonctions définies sur $\mathbb{R}$ (ou $\mathbb{R}^d$). Lorsque cela a un sens, on définit la convolée de f avec g par

$$f \star g \, : \, x \mapsto f \star g(x) = \int_{\mathbb{R}} f(x-t) g(t) dt.$$

On note donc par $\star$ le produit de convolution. La convolée est bien définie lorsque, par exemple, $f \in L^p(\mathbb{R})$ et $g \in L^q(\mathbb{R})$ avec $p, q \geq 1$ tels que $p^{-1} + q^{-1} = 1$. Elle est aussi définie presque partout lorsque $f \in L^1(\mathbb{R})$ et $g \in L^p(\mathbb{R})$. On pourra se référer à Briane et Pagès (2000, Chap. 13) pour des rappels plus détaillés.

2.1 Approximation de l'identité : rappels

L'espace $(L^1(\mathbb{R}), +, ., \star)$ est une algèbre commutative, ne possédant pas d'unité : il n'existe pas de fonction $e \in L^1(\mathbb{R})$ telle que pour toute fonction $f \in L^1(\mathbb{R})$, $f \star e = f$. On pallie cette absence d'unité, en introduisant des suites qui en sont des approximations, au moins asymptotiquement.

Définition 7 *Une famille de fonctions $(g_h)_{h>0} \subset L^1(\mathbb{R})$ est une approximation de l'unité pour la convolution si elle vérifie les trois conditions suivantes :*

(i) pour tout $h > 0$, $\int_{\mathbb{R}} g_h(x)dx = 1$,

(ii) $\sup_{h>0} \int_{\mathbb{R}} |g_h(x)|dx < +\infty$,

(iii) $\forall \varepsilon > 0$, $\lim_{h\to 0} \int_{|x|>\varepsilon} |g_h(x)|dx = 0$.

Proposition 8 *Soit $(g_h)_{h>0} \subset L^1(\mathbb{R})$ une approximation de l'unité. Soit $f \in L^p(\mathbb{R})$, $p \geq 1$. Alors pour tout $h > 0$, $f \star g_h \in L^p(\mathbb{R})$ et*

$$\lim_{h\to 0} \|f \star g_h - f\|_p = 0.$$

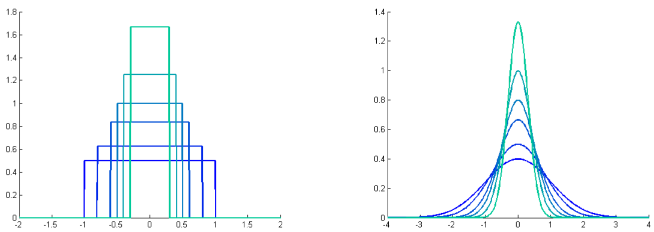
La preuve de ce résultat, ainsi que d'autres résultats d'approximation, peuvent être trouvés au Chapitre 13 du livre de Briane et Pagès (2000) .

2.2 Noyau et approximation

Notation. Dans la suite, pour tout noyau K sur $\mathbb{R}$, et tout réel $h > 0$, on note K_h la fonction, définie sur $\mathbb{R}$ par

$$K_h : x \mapsto \frac{1}{h} K\left(\frac{x}{h}\right). \tag{2}$$

Quelques courbes représentatives sont tracées à la Figure 2.



Noyau K rectangulaire — Noyau K gaussien

FIGURE 2 – Famille de fonctions $(K_h)_{h>0}$.

Proposition 9 *Soit K un noyau, et $(K_h)_{h>0}$ la famille de fonctions associée (voir notation (2)). Alors $(K_h)_{h>0}$ est une approximation de l'unité pour la convolution.*

Exercice. Prouver la Proposition 9.

Conséquence. Les noyaux pourront donc être utilisés en statistique pour leurs propriétés d'approximation. Les deux résultats ci-dessous prouvent que l'on peut même obtenir des résultats plus fins que la Proposition 8 dans le cas où l'on cherche à approcher des fonctions régulières.

Dans la suite, pour $\alpha > 0$, on note par $\lfloor \alpha \rfloor$ le plus grand entier qui est strictement plus petit que α.

2.2.1 Contrôle local de l'erreur d'approximation

On définit les espaces fonctionnels de Hölder, permettant d'évaluer la régularité locale d'une fonction.

Définition 10 *Soient $I \subset \mathbb{R}$ un intervalle, et $\alpha, L \in \mathbb{R}_+^*$. On définit la classe de Hölder $\mathcal{H}_I(\alpha, L)$ comme l'ensemble des fonctions g définies sur I telles que g est $\lfloor\alpha\rfloor-$fois dérivable sur I et telles que, quels que soit $x, x' \in I$,*

$$\left|g^{(\lfloor\alpha\rfloor)}(x) - g^{(\lfloor\alpha\rfloor)}(x')\right| \leq L|x - x'|^{\alpha - \lfloor\alpha\rfloor}. \tag{3}$$

Remarques.

1. Si $\alpha \in]0,1]$, $\lfloor\alpha\rfloor = 0$, on obtient la classe des fonctions contractantes. Si $\alpha = 1$, on obtient les fonctions lipschitziennes.
2. L'indice α d'une classe $\mathcal{H}_I(\alpha, L)$ peut être vu comme l'"indice de régularité" des fonctions de la classe (de la même façon que l'indice k de la classe $\mathcal{C}^k(I)$ des fonctions $k-$fois continûments dérivables sur I indique aussi la régularité de la classe).

Proposition 11 *Soient $\alpha, L \in \mathbb{R}_+^*$. Soit K un noyau d'ordre $l = \lfloor\alpha\rfloor$, tel que $\int_{\mathbb{R}} |u|^\alpha |K(u)| du < +\infty$, et $(K_h)_{h>0}$ la famille de fonctions associée (voir notation (2)). Soit $g \in \mathcal{H}_I(\alpha, L)$. Alors, quels que soient $x_0 \in \mathbb{R}$, $h > 0$,*

$$|K_h \star g(x_0) - g(x_0)| \leq Ch^\alpha, \text{ avec } C = \frac{L}{l!}\int_{\mathbb{R}} |u|^\alpha |K(u)| du.$$

2.2.2 Contrôle global de l'erreur d'approximation

Les espaces de Nikol′skiĭ définis par Nikol′skiĭ (1975) peuvent être compris comme une version intégrée des espaces de Hölder : ils permettent d'évaluer une régularité globale.

Définition 12 *Soient $\alpha, L \in \mathbb{R}_+^*$. On définit la classe de Nikol'skiĭ $\mathcal{N}_2(\alpha, L)$ comme l'ensemble des fonctions g telles que g est $\lfloor\alpha\rfloor$-fois dérivable sur $\mathbb{R}$ et quel que soit $x \in \mathbb{R}$,*

$$\left\|g^{(\lfloor\alpha\rfloor)}(. + x) - g^{(\lfloor\alpha\rfloor)}\right\|_2 = \left(\int_{\mathbb{R}} \left(g^{(\lfloor\alpha\rfloor)}(x' + x) - g^{(\lfloor\alpha\rfloor)}(x')\right)^2 dx'\right)^{1/2} \leq L|x|^{\alpha - \lfloor\alpha\rfloor}.$$

Remarque. On peut plus généralement définir les espaces de Nikol'skiĭ $\mathcal{N}_p(\alpha, L)$ pour n'importe quel $p \geq 1$.

Proposition 13 *Soient $\alpha, L \in \mathbb{R}_+^*$. Soit K un noyau d'ordre $l = \lfloor\alpha\rfloor$, tel que $\int_{\mathbb{R}} |u|^\alpha |K(u)| du < +\infty$, et $(K_h)_{h>0}$ la famille de fonctions associée (voir notation (2)). Soit $g \in \mathcal{N}_2(\alpha, L)$. Alors, quel que soit $h > 0$,*

$$\|K_h \star g - g\|_2 \leq Ch^\alpha, \text{ avec } C = \frac{L}{l!}\int_{\mathbb{R}} |u|^\alpha |K(u)| du.$$

Les Propositions 11 et 13 justifient l'usage des noyaux en statistique : si une fonction g à estimer est suffisamment régulière, on peut la remplacer par $g \star K_h$ pour un certain noyau K. L'intérêt est alors que $g \star K_h$ est une fonction définie par une intégrale, qui se réécrit donc comme une espérance, et un estimateur simple est alors obtenu par la méthode des moments. Le biais de l'estimateur obtenu est alors contrôlé grâce à la Proposition 11 (cas d'un risque ponctuel) ou grâce à la Proposition 13 (cas du risque quadratique intégré).

Références

Marc Briane et Gilles Pagès : *Théorie de l'intégration : licence de mathématiques ; cours et exercices.* Vuibert, 2000.

Fabienne Comte : *Estimation non-paramétrique.* Spartacus IDH, 2015.

Gérard Kerkyacharian, Oleg Lepski et Dominique Picard : Nonlinear estimation in anisotropic multi-index denoising. *Probab. Theory Related Fields*, 121(2):137–170, 2001.

Gilles Lacombe et Pascal Massat : *Analyse fonctionnelle : exercices corrigés ;[2e cycle universitaire-agrégation].* Dunod, 1999.

Sergeĭ Mikhaĭlovich Nikol′skiĭ : *Approximation of functions of several variables and imbedding theorems.* Springer-Verlag, New York, 1975. Translated from the Russian by John M. Danskin, Jr., Die Grundlehren der Mathematischen Wissenschaften, Band 205.

Alexandre B. Tsybakov : *Introduction to nonparametric estimation.* Springer Series in Statistics. Springer, New York, 2009. Revised and extended from the 2004 French original, Translated by Vladimir Zaiats.