

REVUE DE STATISTIQUE APPLIQUÉE

MICHEL LEJEUNE

**Estimation non-paramétrique par noyaux :
régression polynomiale mobile**

Revue de statistique appliquée, tome 33, n° 3 (1985), p. 43-67

http://www.numdam.org/item?id=RSA_1985__33_3_43_0

© Société française de statistique, 1985, tous droits réservés.

L'accès aux archives de la revue « *Revue de statistique appliquée* » (<http://www.sfds.asso.fr/publicat/rsa.htm>) implique l'accord avec les conditions générales d'utilisation (<http://www.numdam.org/conditions>). Toute utilisation commerciale ou impression systématique est constitutive d'une infraction pénale. Toute copie ou impression de ce fichier doit contenir la présente mention de copyright.

NUMDAM

Article numérisé dans le cadre du programme
Numérisation de documents anciens mathématiques
<http://www.numdam.org/>

ESTIMATION NON-PARAMETRIQUE PAR NOYAUX : REGRESSION POLYNOMIALE MOBILE

Michel LEJEUNE
Université de Lausanne

RESUME

Pour un estimateur non-paramétrique à noyau d'une courbe de régression l'optimalité du noyau est étudiée dans le cas d'échantillonnage fini avec abscisses équiréparties. On démontre que, dans la classe des noyaux éliminant le biais jusqu'à un degré k d'approximation polynomiale locale de la courbe, le noyau donnant la variance minimale se déduit du polynôme de Legendre de degré k , ou éventuellement de degré $k + 1$ si l'on impose la continuité aux extrémités du support. Le cas des effets de bord est aussi envisagé : les noyaux optimaux sont également déterminés.

On démontre alors qu'il y a équivalence entre l'approche par noyaux optimaux et une solution par régression polynomiale mobile avec un choix de pondérations adéquat. Cette dernière solution permet une généralisation à des abscisses de répartition quelconque avec ou sans effets de bord. Quelques exemples de simulations sont donnés.

ABSTRACT

Optimal kernels are studied for non parametric regression in the case of finite sampling with equidistant x values. Within the class of kernels which yield unbiasedness up to a local polynomial approximation of order k for the regression curve, it is shown that the minimum variance kernel can be derived from Legendre polynomials of degree k or, possibly, of degree $k + 1$ when imposing continuity on both ends of the support. Boundary effects are also considered and corresponding kernels are exhibited.

It is then shown that the optimal kernel approach and a solution through polynomial regression with appropriate weights are equivalent. This later solution can be generalized to any distribution of x values with or without boundary effects. Some simulations are given.

1. INTRODUCTION

Soit $\{Y(x)\}$ une famille de variables aléatoires indexées par $x \in [a, b]$ et telle que pour tout $x \in [a, b]$:

$$E[Y(x)] = \mu_Y(x)$$

existe et :

$$\text{Var}[Y(x)] = \sigma^2$$

Soit $\{x_i; i = 1, \dots, n\}$ une suite de points dans $\mathbf{R}$ ordonnés et équirépartis à l'intérieur de $[a, b]$. Par commodité nous poserons $y_i = Y(x_i)$. Le problème envisagé est l'estimation de $\mu_Y(x)$ en tout point x de $[a, b]$ avec comme seule hypo-

thèse que μ_Y est une fonction indéfiniment dérivable et analytique sur cet intervalle en ce sens que son développement en série de Taylor au voisinage de x quelconque dans $[a, b]$ converge vers $\mu_Y(x)$. Le problème est non-paramétrique par le fait que μ_Y n'appartient pas, au contraire de l'estimation paramétrique classique, à une famille de fonctions décrites par un nombre fini de paramètres réels.

Notons que le modèle posé n'est pas, stricto sensu, celui de la régression puisqu'ici les x_i sont fixés et non aléatoires. Ce problème général sera considéré dans la section 6.

Considérons la classe des estimateurs linéaires, par rapport aux y_i :

$$\hat{\mu}_Y(x) = \frac{\sum w_i y_i}{\sum w_i} \quad (2.1)$$

où le signe $\sum$ s'entend, ici ainsi que par la suite sauf mention contraire, comme une sommation de 1 à n , et où w_i est fonction de x et de x_i .

Lorsque les pondérations w_i peuvent s'écrire sous la forme :

$$w_i = \begin{cases} \frac{1}{\delta} K\left(\frac{x_i - x}{\delta}\right) & \text{si } |x_i - x| \leq \delta \\ 0 & \text{sinon} \end{cases}$$

où K est une fonction continue sur $[-1, +1]$ vérifiant

$$\int_{-1}^1 K(u) du = 1$$

nous dirons que nous avons un estimateur à noyau, K étant précisément le noyau. L'intervalle $[x - \delta, x + \delta]$ est appelé fenêtre en x et 2δ est appelé largeur de fenêtre. Notons que δ est un facteur d'échelle dont le choix est essentiel. Nous n'envisagerons toutefois pas ce choix ici et raisonnerons à valeur de δ fixée pour nous intéresser spécifiquement à la forme du noyau caractérisée par la fonction K . D'autre part nous excluons d'emblée les noyaux dont le support n'est pas borné.

L'idée générale prévaut en estimation non-paramétrique par noyaux que la forme du noyau est secondaire en regard de la détermination de la largeur de fenêtre. Ceci est vrai si l'on se restreint, comme on le fait généralement, à la classe des noyaux positifs et si l'on ne se préoccupe que de propriétés asymptotiques. Dans la pratique rien ne sert de viser une largeur de fenêtre asymptotiquement optimale si le noyau n'évite pas tant soit peu les fortes distorsions dues au biais. Les précurseurs de l'estimation non-paramétrique qui étudiaient le filtrage des séries chronologiques ne se sont pas trompés sur la priorité à accorder entre largeur et forme du noyau puisqu'ils ont essentiellement proposé des noyaux non positifs de largeur fixe.

Dans notre étude nous verrons en quoi les noyaux non nécessairement positifs améliorent sensiblement le biais dans le cas de la régression non-paramétrique. Nous mettrons en évidence une méthode de lissage qui conduit à l'optimalité du noyau pour la situation d'équirépartition des abscisses et même pour la situation plus générale, considérée à la section 6, d'une répartition quelconque. De plus, les effets de bords intervenant aux extrémités du domaine des abscisses sont parfaitement neutralisés.

L'intérêt de cette étude réside dans le fait qu'on considère un échantillonnage fini alors que les travaux dans ce domaine ont principalement porté jusqu'à présent sur des considérations asymptotiques reposant en particulier sur une largeur de fenêtre tendant vers zéro. De tels résultats ont un intérêt limité car la largeur de fenêtre nécessaire à un lissage correct reste importante même pour de grands échantillons. On pourra consulter à ce sujet les résultats de simulations effectuées par DODGE et LEJEUNE (1982).

2. BIAIS ET VARIANCE (CAS D'ABSCISSES EQUIREPARTIES)

L'approche que nous adoptons s'inspire des travaux de GASSER et MULLER (1979, 1982) qui se fondent, pour $n \rightarrow \infty$, $\delta \rightarrow 0$ et $n\delta \rightarrow \infty$, sur les résultats asymptotiques suivants :

$$E [\hat{\mu}_y(x) - \mu_y(x)] = \int_{-1}^1 K(u) [\mu_y(x + \delta u) - \mu_y(x)] du + O\left(\frac{1}{n\delta} + \frac{1}{n^{\alpha-1}}\right) \quad (2.2)$$

$$\text{Var} [\hat{\mu}_y(x)] = \frac{\sigma^2}{n\delta} \int_{-1}^1 K^2(u) du + O\left(\frac{1}{n^2 \delta^2} + \frac{1}{n^{\alpha} \delta}\right) \quad (2.3)$$

où les termes en n^{α} et $n^{\alpha-1}$ résultent du fait que les auteurs n'imposent aux x_i que l'équidistance asymptotique avec :

$$\max_i |x_i - x_{i-1}| = O\left(\frac{1}{n^{\alpha}}\right).$$

Ces formules sont écrites pour une répartition des abscisses sur $[0, 1]$, la variance devant être multipliée par $b-a$ si la répartition s'étend sur $[a, b]$. Elles ne valent que lorsque x n'induit pas des "effets de bord" c'est-à-dire lorsque $x \in [a + \delta, b - \delta]$. Toutefois GASSER et MULLER donnent les formules dans le cas d'effets de bord qui, essentiellement, ne diffèrent des précédentes que par les ordres d'approximation.

Montrons que les formules 2.2 et 2.3 peuvent être établies sans difficulté pour un échantillonnage fini avec une approximation dont nous précisons la nature, ce qui nous permettra dès lors d'envisager la recherche d'optimalité. Nous commençons par établir un lemme ayant trait à l'approximation intégrale de sommes discrètes.

Lemme

Soit f une fonction réelle, continue et indéfiniment dérivable sur $[a, b]$ telle que $f'(a) = f'(b)$. Pour une suite croissante de points $\{x_1, x_2, \dots, x_n\}$ telle que $\Delta = x_{i+1} - x_i$ soit constant et $a = x_1 - \frac{\Delta}{2}$, $b = x_n + \frac{\Delta}{2}$, on a :

$$\sum_{i=1}^n f(x_i) = \frac{n}{b-a} \int_a^b f(x) dx + O\left(\frac{1}{n^3}\right). \quad (2.4)$$

S'il existe un point (ou tout au plus un nombre fini de points) de (a, b) où f n'est pas continue mais finie alors l'approximation vaut avec $O(1)$.

Démonstration

Nous prenons d'abord f dérivable partout sur $[a, b]$. En développant f en série de Taylor au voisinage de x_i puis en intégrant sur $\left[x_i - \frac{\Delta}{2}, x_i + \frac{\Delta}{2}\right]$ et en sommant sur i , on obtient aisément :

$$\sum_{i=1}^n f(x_i) = \frac{1}{\Delta} \int_a^b f(x) dx - \frac{\Delta^2}{24} \sum f''(x_i) + O(\Delta^3). \quad (2.5)$$

Nous pouvons appliquer cette formule pour f'' et utiliser la propriété $f'(a) = f'(b)$, ce qui donne :

$$\sum_{i=1}^n f''(x_i) = -\frac{\Delta^2}{24} \sum_{i=1}^n f^{(4)}(x_i) + O(\Delta^3)$$

soit en substituant dans (2.5) :

$$\sum_{i=1}^n f(x_i) = \frac{1}{\Delta} \int_a^b f(x) dx + O(\Delta^3)$$

ce qui est équivalent à (2.4) puisque $\Delta = (b - a)/n$.

Soit maintenant un point de discontinuité situé à l'intérieur de $x_K \pm \frac{\Delta}{2}$ (le cas d'une borne ne posant aucun problème particulier). On peut écrire :

$$\begin{aligned} \sum_{i=1}^n f(x_i) &= \frac{1}{\Delta} \int_a^b f(x) dx - \frac{1}{\Delta} \int_{x_K - \frac{\Delta}{2}}^{x_K + \frac{\Delta}{2}} f(x) dx + f(x_K) + O(\Delta^3) \\ &= \frac{1}{\Delta} \int_a^b f(x) dx + \epsilon_K + O(\Delta^3) \end{aligned}$$

avec $|\epsilon_K| \leq M_K - f(x_K)$ où M_K est le maximum de f sur $\left[x_K - \frac{\Delta}{2}, x_K + \frac{\Delta}{2}\right]$, donc $\epsilon_K = O(1)$. Notons que si f est continue mais non dérivable en x_K l'approximation est $o(1)$.

Calculons maintenant biais et variance pour l'estimateur (2.1) :

$$\text{Biais} [\hat{\mu}_y(x)] = E [\hat{\mu}_y(x) - \mu_y(x)] = \frac{\sum w_i [\mu_y(x_i) - \mu_y(x)]}{\sum w_i} \quad (2.6)$$

et

$$\text{Var} [\mu_y(x)] = E [\hat{\mu}_y(x) - E\hat{\mu}_y(x)]^2 = \sigma^2 \frac{\sum w_i^2}{(\sum w_i)^2}. \quad (2.7)$$

Supposons que μ_y soit suffisamment dérivable sur le domaine d'intérêt, que les w_i soient donnés par un noyau K indéfiniment dérivable sur $[-1, +1]$ et tel que $K(\pm 1) = K'(\pm 1) = 0$. Supposons encore que le point x n'induisse pas des effets de bord c'est-à-dire que $x_1 + \delta \leq x \leq x_n - \delta$, où δ est la largeur

de fenêtre. Dorénavant un tel point sera appelé un "point intérieur" par opposition à un point du bord gauche ou droit. Alors le lemme s'applique immédiatement à la formule (2.6) du biais (en définissant a et b opportunément) :

$$\text{Biais } [\hat{\mu}_y(x)] = \frac{\frac{n}{b-a} \int_{x-\delta}^{x+\delta} \frac{1}{\delta} K\left(\frac{u-x}{\delta}\right) [\mu_y(u) - \mu_y(x)] du + o\left(\frac{1}{n^3}\right)}{\frac{n}{b-a} \int_{x-\delta}^{x+\delta} \frac{1}{\delta} K\left(\frac{u-x}{\delta}\right) du + o\left(\frac{1}{n^3}\right)}$$

et puisque K est normalisé sur $[-1, +1]$:

$$\begin{aligned} \text{Biais } [\hat{\mu}_y(x)] &= \int_{x-\delta}^{x+\delta} \frac{1}{\delta} K\left(\frac{u-x}{\delta}\right) [\mu_y(u) - \mu_y(x)] du + o\left(\frac{1}{n^3}\right) \\ \text{Biais } [\hat{\mu}_y(x)] &= \int_{-1}^1 K(u) \cdot [\mu_y(x + \delta u) - \mu_y(x)] du + o\left(\frac{1}{n^4}\right) \quad (2.8) \end{aligned}$$

Dans le cas où on utilise un noyau avec des discontinuités – respectivement non dérivabilités – aux bornes du support, cette formule ne vaut qu'avec $o\left(\frac{1}{n}\right)$ – respectivement $o\left(\frac{1}{n}\right)$.

Supposons maintenant que x soit un point de bord, par exemple à gauche, et donc que $x = x_1 + q\delta$ avec $0 \leq q < 1$. Dans ce cas nous utiliserons par commodité pour la recherche d'optimalité qui suivra une troncature du noyau K telle que $K(u) = 0$ pour $u < -q$. Nous sommes alors ramenés au cas d'une discontinuité comme précédemment.

Pour la variance on a :

$$\begin{aligned} \text{Var } [\hat{\mu}_y(x)] &= \sigma^2 \frac{\frac{n}{b-a} \int_{x-\delta}^{x+\delta} \frac{1}{\delta^2} K^2\left(\frac{u-x}{\delta}\right) du + o\left(\frac{1}{n^3}\right)}{\left[\frac{n}{b-a} \int_{x-\delta}^{x+\delta} \frac{1}{\delta} K\left(\frac{u-x}{\delta}\right) du + o\left(\frac{1}{n^3}\right) \right]^2} \\ \text{Var } [\hat{\mu}_y(x)] &= \frac{b-a}{n\delta} \sigma^2 \int_{-1}^{+1} K^2(u) du + o\left(\frac{1}{n^5}\right) \quad (2.9) \end{aligned}$$

Dans le cas du noyau discontinu ou d'effets de bord l'approximation vaut avec $o\left(\frac{1}{n^2}\right)$

Ces résultats amènent quelques remarques. Tout d'abord les formules (2.8) et (2.9) sont les formules (2.2) et (2.3) dans le cas d'une discontinuité aux bornes du support, si l'on omet les termes relatifs à la condition de quasi-équidistance des abscisses, à ceci près que δ n'intervient pas dans les ordres d'approximation puisque nous nous plaçons dans l'hypothèse d'un δ fixe indépendant de n . On notera avec intérêt que, δ étant fixé, la nature des termes négligés tient à l'approximation continue, par intégrale, d'une somme discrète de termes. Ces formules s'appliquent sans autre hypothèse aux points situés sur les bords, la seule modification portant sur l'introduction d'une discontinuité du noyau due à sa troncature.

3. OPTIMISATION DU NOYAU

Par optimisation nous entendons la réduction la plus forte possible des deux types d'erreurs d'estimation que sont le biais et la variance. Traditionnellement on utilise comme critère d'optimisation l'erreur quadratique moyenne qui combine ces deux types. Nous préférons considérer séparément l'étude du biais et de la variance pour nous situer dans une plus grande généralité et rapprocherons ensuite nos résultats de ceux rencontrés dans la littérature.

Pour arriver à l'optimum nous devons choisir deux entités indépendamment ou non : d'une part la forme du noyau, c'est-à-dire la fonction K de support $[-1, +1]$ et d'autre part la largeur de fenêtre δ . Partant de l'idée que le choix de δ sera en pratique guidé par des considérations descriptives se rapportant au lissage de la courbe, nous allons supposer que δ est fixé à une valeur qui peut être optimale ou non et nous intéresser essentiellement à l'allure la plus bénéfique du noyau K quel que soit δ .

En considérant les formules d'approximation du biais et de la variance (2.8) et (2.9), on constate que la minimisation du biais est plus problématique puisqu'elle va dépendre de l'allure de la vraie courbe de régression μ_y entre $x - \delta$ et $x + \delta$. Par contre le problème de la variance est plus simple car il s'agit de minimiser uniquement :

$$\int_{-1}^{+1} K^2(u) du. \quad (3.10)$$

Aussi adopterons-nous la démarche suivante. Dans un premier temps, nous allons définir pour le noyau les contraintes qui permettent de contrôler le biais jusqu'à un niveau voulu. Plus précisément nous écrirons les conditions pour lesquelles le noyau annule le biais pour une approximation polynômiale de Taylor de μ_y au voisinage de x de degré donné. A la suite de quoi nous rechercherons, dans l'ensemble des noyaux remplissant ces conditions, celui qui minimise (3.10). Cela signifie que nous privilégions le biais par rapport à la variance (et implicitement les aspects descriptifs par rapport aux aspects inférentiels) mais sans que cela soit réellement préjudiciable car la variance peut être aussi contrôlée par la largeur de fenêtre qui apparaît à son dénominateur. Nous envisagerons d'abord le cas d'un point x intérieur au domaine, le cas d'un point de bord faisant l'objet de la section suivante.

Définition

On dit qu'un noyau K est d'ordre r si ses moments sont nuls jusqu'à l'ordre $r - 1$ et son moment d'ordre r est non nul, soit :

$$\int_{-1}^1 u^K K(u) du = \begin{cases} 0 & \text{si } K = 1, \dots, r - 1 \\ m_r \neq 0 & \text{si } K = r. \end{cases}$$

Considérons le développement en série de TAYLOR ;

$$\mu_y(x + \delta u) = \mu_y(x) + \sum_{K=1}^{\infty} (\delta u)^K \mu_y^{(K)}(x) / K!$$

En substituant dans l'approximation (2.8) du biais on obtient :

$$\text{Biais } [\hat{\mu}_Y(x)] = \sum_{K=1}^{\infty} \frac{\delta^K}{K!} \mu^{(K)}(x) \int_{-1}^1 u^K K(u) du.$$

Pour un noyau K_r d'ordre r les $r-1$ premiers termes de cette somme sont nuls. Soit $R_r(\delta u)$ la sommation à partir de $K=r$ dans le développement de $\mu_Y(x+\delta u)$, le biais vaut :

$$\int_{-1}^1 K_r(u) R_r(\delta u) du.$$

Quand $r \rightarrow \infty$ le reste R_r converge vers zéro uniformément en δu , mais ce n'est que pour une suite de noyaux K_r appropriée que le biais tendra vers zéro. Néanmoins, pour r fixé, le biais est éliminé pour l'approximation locale de μ_Y par le polynôme de Taylor de degré $r-1$. Par ailleurs la formule ci-dessus montre que la réduction du biais dépendra de la précision de cette approximation de μ_Y dans la fenêtre $[x-\delta, x+\delta]$.

D'un point de vue descriptif on imagine l'intérêt qu'il y aura à choisir un noyau d'ordre élevé si l'on veut éviter de forts biais dans les zones de μ_Y à forte variation de courbure. Notons par exemple que les noyaux non négatifs que l'on utilise généralement ne peuvent avoir un deuxième moment qui s'annule et, par conséquent, vont donner un décalage systématique dès qu'il y aura courbure.

Montrons que dans la classe des noyaux d'ordre r pair il existe un et un seul noyau de forme polynomiale de degré $r-2$. On peut se limiter à l'étude de r pair car, en raison de la symétrie du problème autour du point x , le noyau doit être une fonction paire et les moments d'ordre impair sont nécessairement nuls. Pour un polynôme de degré $r-2$, où r est pair, les contraintes portant sur les $r-1$ premiers moments, ainsi que la contrainte de normalisation se traduisent par :

$$\begin{pmatrix} 1 & 1/3 & 1/5 & \dots & 1/(r-1) \\ 1/3 & 1/5 & 1/6 & \dots & 1/(r+1) \\ \vdots & & & & \vdots \\ \vdots & & & & \vdots \\ 1/(r-1) & & \dots & & 1/(2r-3) \end{pmatrix} \tilde{\alpha} = \begin{pmatrix} 1/2 \\ 0 \\ \vdots \\ \vdots \\ 0 \end{pmatrix} \quad (3.11)$$

où $\tilde{\alpha}$ est le vecteur des coefficients pairs du polynôme. La matrice ci-dessus étant régulière, le système admet bien une solution unique pour $\tilde{\alpha}$. A un facteur de proportionnalité près cette solution est donnée par les cofacteurs de la première ligne. On trouve pour les premières valeurs de r (voir également la figure 1) les noyaux polynomiaux optimaux ϕ_r suivants :

$$\begin{aligned} r=2 \quad \phi_2(x) &= \frac{1}{2} \\ r=4 \quad \phi_4(x) &= \frac{3}{8} (3 - 5x^2) \\ r=6 \quad \phi_6(x) &= \frac{15}{128} (15 - 70x^2 + 63x^4) \end{aligned} \quad (3.12)$$

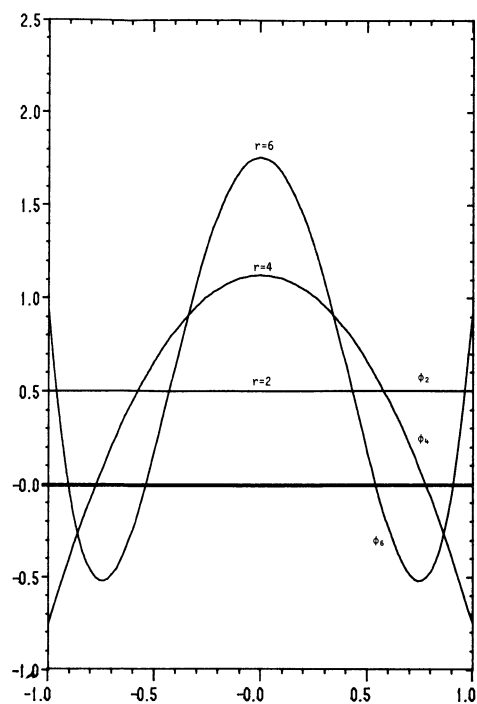


Figure 1. — Noyaux optimaux d'ordres $r = 2, 4$ et 6 .

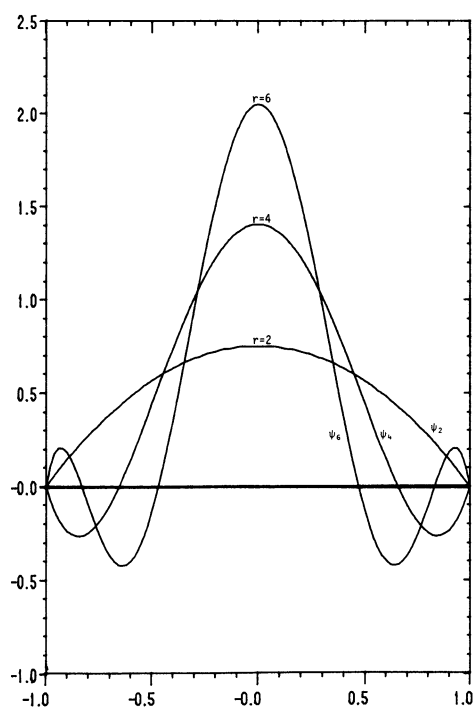


Figure 2. — Noyaux continus aux bornes optimaux d'ordres $r = 2, 4$ et 6 .

Ces noyaux qui sont les polynômes de degré minimal pour chaque valeur de r présentent une discontinuité aux bornes du support ce qui entraîne une discontinuité pour l'estimation lorsqu'on déplace l'abscisse. On peut donc préférer choisir un polynôme vérifiant une contrainte de continuité supplémentaire pour (3.11) :

$$(1 \quad 1 \dots 1) \tilde{\alpha} = 0,$$

ce qui oblige à considérer un polynôme de degré r au lieu de $r - 2$. On trouve alors pour les premières valeurs de r (voir également la figure 2) les noyaux polynômiaux continus aux bornes et optimaux d'ordre r , notés ψ_r et définis par :

$$\begin{aligned} r = 2 \quad \psi_2(x) &= \frac{3}{4}(1 - x^2) \\ r = 4 \quad \psi_4(x) &= \frac{15}{32}(1 - x^2)(3 - 7x^2) \\ r = 6 \quad \psi_6(x) &= \frac{105}{256}(1 - x^2)(5 - 30x^2 + 33x^4) \end{aligned} \quad (3.13)$$

De même on peut introduire encore une contrainte de dérivabilité aux bornes du support :

$$(0 \quad 2 \quad 4 \dots r + 2) \tilde{\alpha} = 0$$

ceci au détriment de l'élévation au degré $r + 2$ pour le polynôme de degré minimal. Pour $r = 2$ on trouve ainsi le noyau biquadratique :

$$\frac{15}{16}(1 - x^2)^2.$$

Venons maintenant à la minimisation de la variance c'est-à-dire de l'expression (3.10) et ne considérons pour le moment que les noyaux de type polynômial. Pour K polynôme de degré m la minimisation de $\int K^2$ implique les conditions :

$$\int_{-1}^1 K \frac{\delta K}{\delta \alpha_i} = 0 \quad i = 1, \dots, m$$

ou encore :

$$\int_{-1}^1 x^i K = 0 \quad i = 1, \dots, m$$

Ces conditions étant identiques à celles exigées pour l'annulation du biais, nous pouvons en déduire que les polynômes de degré $r - 2$ mis en évidence précédemment — voir (3.12) — sont également de variance minimale parmi l'ensemble des polynômes de degré $r - 2$. En fait on peut énoncer un résultat beaucoup plus général.

Proposition

Parmi les noyaux K d'ordre r (nécessairement pair) le noyau proportionnel à $\frac{1}{x} P_{r-1}(x)$, où P_{r-1} est le polynôme de Legendre de degré $r - 1$, est celui qui minimise $\int K^2$.

La démonstration est un cas particulier de résultats plus vastes donnés par MAMMITZSCH (1983). Rappelons que les polynômes de Legendre peuvent être définis par l'équation de récurrence :

$$(n+1)P_{n+1}(x) = (2n+1)xP_n(x) - nP_{n-1}(x) \quad \text{avec} \quad P_0(x) = 1, P_1(x) = x$$

et qu'ils vérifient :

$$\forall n, \forall m, \int_{-1}^1 P_n(x) P_m(x) dx = 0 \quad \text{et} \quad \int_{-1}^1 P_n^2(x) dx = \frac{2}{2n+1}$$

Posons donc :

$$Q_n = \sqrt{\frac{2n+1}{2}} P_n$$

alors $Q_0, Q_1, Q_2, \dots$ forment une famille orthonormée de fonctions qui est dense dans l'ensemble $L^2[-1, +1]$. Pour toute fonction dans cet ensemble les conditions relatives aux $r-1$ premiers moments s'écrivent avec les notations usuelles du produit intérieur dans L^2 :

$$\langle K, x^k \rangle = \begin{cases} 1 & \text{si } k = 0 \\ 0 & \text{si } k = 1, \dots, r-1 \end{cases}$$

ou de façon équivalente :

$$\langle K, Q_m \rangle = q_m^0 \quad m = 0, 1, \dots, r-1$$

où q_m^0 est le terme constant de Q_m . Le noyau K s'écrit alors :

$$K = q_0^0 Q_0 + q_1^0 Q_1 + \dots + q_{r-1}^0 Q_{r-1} + \sum_{k=r}^{\infty} a_k Q_k,$$

d'où :

$$\langle K, K \rangle = (q_0^0)^2 + (q_1^0)^2 + \dots + (q_{r-1}^0)^2 + \sum_{k=r}^{\infty} a_k^2$$

et la minimisation de cette expression est obtenue lorsque les a_k sont tous nuls. (Remarquons que q_m^0 étant nul pour m impair, on retrouve le fait que r est nécessairement pair). En utilisant les propriétés récursives des polynômes de Legendre,

on montre qu'il s'agit, à un coefficient de proportionnalité près, de $\frac{1}{x} P_{r-1}(x)$.

Notons que ce polynôme a $r-2$ racines réelles dans $[-1, +1]$.

A ce point il est intéressant d'étudier comment évolue la variance minimale atteignable pour différentes valeurs de r . Pour $r = 2, 4, 6$, on trouve respectivement une valeur minimale de $\int K^2$ égale à 0.500, 1.125, 1.758. On voit donc que la variance se dégrade assez nettement lorsque l'ordre s'élève et qu'il faudra limiter nos ambitions dans la réduction du biais à une valeur de r pas trop élevée. On démontre d'ailleurs que la variance diverge quand $r \rightarrow \infty$.

Si nous imposons maintenant que le noyau soit continu aux bornes du support à savoir $K(\pm 1) = 0$, la minimisation devient plus délicate et nous ne dispo-

sons que de résultats partiels. Dans son article devenu classique et relatif à l'estimation de densité où un problème d'optimisation similaire est posé, EPANECHNIKOV (1969) a montré que le noyau $\frac{3}{4}(1-x^2)$ minimise $\int K^2$ parmi les noyaux dont le deuxième moment est égal à 1. En fait GASSER et MULLER (1979) ont montré, pour $r = 2$ et $r = 4$, que les solutions ψ_r données en (3.13) minimisent $\int K^2$ parmi les noyaux continus aux bornes et ayant un nombre de changements de signe minimal (à savoir égal à $r-2$) sur $(-1, 1)$. Avec GASSER et MULLER nous conjecturons que ceci se généralise à $r > 4$.

Conjecture

Parmi les noyaux d'ordre r (pair) montrant un nombre de changements de signe minimal (égal à $r-2$) sur $(-1, +1)$ et s'annulant en ± 1 , le noyau proportionnel à $\frac{1}{x} \int_0^x P_r(u) du$, où P_r est le polynôme de Legendre de degré r , est celui qui minimise $\int K^2$.

Le calcul de $\int K^2$ donne pour $r = 2, 4, 6$ respectivement 0.600, 1.250, 1.893. En comparant aux valeurs obtenues avec les noyaux quelconques aux bornes du support, on voit que l'augmentation de la variance n'est pas significative en regard du gain que l'on peut attendre sur le plan du lissage. On peut pousser la comparaison plus loin en regardant la valeur du r -ième moment m_r qui constitue, à une constante multiplicative près, le terme principal du biais pour δ petit :

$$\int_{-1}^1 \mu_Y^{(r)}(x) \frac{(\delta u)^r}{r!} K(u) du = \mu_Y^{(r)}(x) \cdot \frac{\delta^r m_r}{r!}$$

dans son développement en série de Taylor. On trouve pour valeurs de m_r :

	$r = 2$	$r = 4$	$r = 6$
"Noyau optimal discontinu"	0.333	- 0.086	0.022
"Noyau optimal continu"	0.200	- 0.048	0.012

A ordre égal on peut donc s'attendre à ce qu'en plus d'un meilleur aspect descriptif le noyau optimal continu présente une amélioration sensible du biais (approximativement réduit de moitié). Globalement le noyau continu peut être jugé préférable mais en notant bien que son degré est r au lieu de $r-2$ pour le noyau discontinu, ce qui peut influencer sur la facilité des calculs. Le choix que nous proposerions à ce point serait celui du noyau continu avec $r = 4$ qui reste acceptable sur le plan de la variance et de la simplicité, et donne un biais nul pour une approximation locale de la courbe de régression par un polynôme du troisième degré, ce qui est tout à fait suffisant en pratique.

Comment ces résultats se rapprochent-ils des résultats classiques rencontrés dans la littérature ? Que ce soit pour l'estimation d'une fonction de densité ou d'une fonction de régression, la procédure usuelle consiste à développer asymptotiquement l'erreur quadratique moyenne (e.q.m.) en un point x pour un estimateur à noyau K de largeur δ . On détermine alors la valeur de δ globalement optimale c'est-à-dire qui minimise l'e.q.m. intégrée (e.q.m.i.) le long de l'intervalle de variation de x . On substitue alors cette valeur dans l'e.q.m.i. qui devient proportionnelle à :

$$n^{-\frac{2r}{2r+1}} [|m_r| \cdot (\int K^2)^r]^{-\frac{2}{2r+1}} \quad (3.14)$$

pour un noyau d'ordre r et de r -ième moment égal à m_r . L'avantage généralement invoqué pour le choix de $r > 2$ est d'améliorer la vitesse de décroissance de l'erreur en fonction de n . Pour les développements détaillés nous renvoyons à TAPIA et THOMPSON (1980) pour l'estimation de densité et GASSER et MULLER (1979) pour la régression : le choix de K qui s'impose alors est celui qui minimise :

$$|m_r|^{\frac{1}{r}} \cdot \int K^2 \quad \text{avec} \quad m_r = \int_{-1}^1 u^r K(u) du$$

Notre propre approche présente l'avantage de reposer sur des notions d'optimalité définies plus simplement et d'explicitier le choix à opérer entre les divers noyaux, tout en s'en tenant à un nombre minimal de conditions et à la seule approximation d'une somme discrète par une intégrale.

Avant de nous prononcer sur le choix d'un noyau nous devons étudier les problèmes dus aux effets de bord qui, dans la pratique, peuvent concerner la moitié du domaine des abscisses, la valeur de δ couvrant couramment le quart de ce domaine.

4. EFFETS DE BORD

Sur le fond le problème dans le cas d'un point x sur le bord gauche ou droit ne diffère pas du problème précédent même s'il est plus difficile d'obtenir des solutions explicites. En effet, comme nous l'avons indiqué plus haut les formules (2.8) et (2.9) du biais et de la variance s'appliquent de la même façon à cette nuance près que le noyau est tronqué sur un côté. Dans ce qui suit nous traiterons le cas du bord gauche, celui du bord droit se réglant symétriquement en changeant x en $-x$.

Soit $x = x_1 + q\delta$ où x_1 est le minimum des x_i et $0 \leq q < 1$, alors le support de K est $[-q, 1]$. Nous pouvons suivre la même démarche qu'à la section précédente consistant tout d'abord à mettre en évidence le noyau unique d'ordre r de forme polynômiale de degré minimal, à ceci près que nous n'avons plus la propriété de symétrie, donc que r peut être pair ou impair, et que vont intervenir dans l'écriture des moments des éléments :

$$\rho_i = \int_{-q}^1 x^i dx = \frac{1 + (-1)^i q^{i+1}}{i+1} \quad (4.15)$$

Un noyau d'ordre r devra être de degré au moins égal à $r-1$ et les équations (3.11) deviennent :

$$\begin{pmatrix} \rho_0 & \rho_1 & \rho_2 & \cdots & \rho_{r-1} \\ \rho_1 & \rho_2 & \rho_3 & \cdots & \rho_r \\ \vdots & & & & \vdots \\ \rho_{r-1} & \cdots & \rho_{2r-2} \end{pmatrix} \tilde{\alpha} = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \quad (4.16)$$

Donnons, à un coefficient de proportionnalité près, les noyaux pour les premiers ordres :

$$r = 1 \quad 1$$

$$r = 2 \quad 2(1 + q^3) - 3(1 - q^2)x$$

$$r = 3 \quad 48(1 + q^3)(1 + q^5) - 45(1 - q^4)^2 + x[60(1 + q^3)(1 - q^4) - 72(1 - q^2)(1 + q^5)] + x^2[90(1 - q^2)(1 - q^4) - 80(1 + q^3)^2].$$

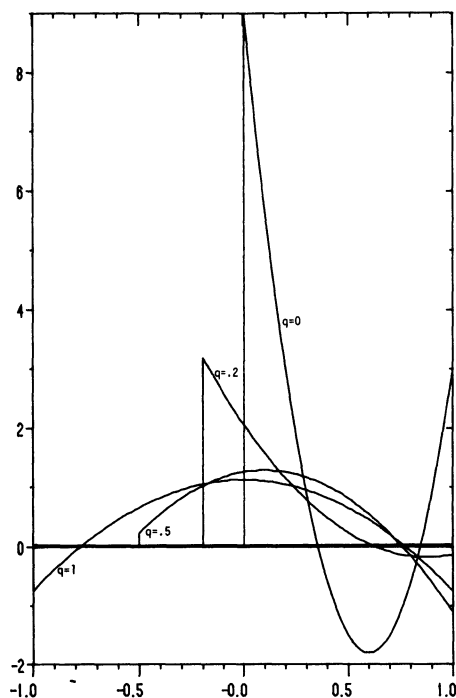


Figure 3. — Effets de bord : noyaux optimaux d'ordre 3 de supports $[-q, 1]$.

La figure 3 donne pour quelques valeurs de q l'allure du noyau d'ordre 3. Notons que pour $q = 1$ on retrouve les noyaux optimaux de la section précédente. Etant donné que nous n'avons plus le support $[-1, +1]$ la solution de variance minimale pour un ordre donné est difficile à obtenir. Nous nous contenterons d'arguer de la continuité existant entre le cas d'effets de bord et le cas sans effets de bord pour voir dans les solutions du type (4.16) une certaine optimalité.

Nous pouvons également ici ajouter une contrainte de continuité mais qui n'est utile que pour la valeur 1. Le noyau polynômial devra alors être au moins de degré r . Par exemple pour $r = 3$ nous devons résoudre le système :

$$\begin{pmatrix} \rho_0 & \rho_1 & \rho_2 & \rho_3 \\ \rho_1 & \rho_2 & \rho_3 & \rho_4 \\ \rho_2 & \rho_3 & \rho_4 & \rho_5 \\ 1 & 1 & 1 & 1 \end{pmatrix} \tilde{\alpha} = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}$$

A un coefficient de proportionnalité près la solution est donnée par les cofacteurs de la première ligne, soit pour α_0 par exemple :

$$\alpha_0 = \begin{vmatrix} \rho_2 & \rho_3 & \rho_4 \\ \rho_3 & \rho_4 & \rho_5 \\ 1 & 1 & 1 \end{vmatrix} = \begin{vmatrix} \rho_2 - \rho_3 & \rho_3 - \rho_4 \\ \rho_3 - \rho_4 & \rho_4 - \rho_5 \end{vmatrix} \quad (4.17)$$

Et de même on obtient

$$\alpha_1 = \begin{vmatrix} \rho_1 - \rho_3 & \rho_3 - \rho_4 \\ \rho_2 - \rho_4 & \rho_4 - \rho_5 \end{vmatrix}, \alpha_2 = \begin{vmatrix} \rho_1 - \rho_2 & \rho_2 - \rho_4 \\ \rho_2 - \rho_3 & \rho_3 - \rho_5 \end{vmatrix}, \alpha_3 = \begin{vmatrix} \rho_1 - \rho_2 & \rho_2 - \rho_3 \\ \rho_2 - \rho_3 & \rho_3 - \rho_4 \end{vmatrix} \quad (4.18)$$

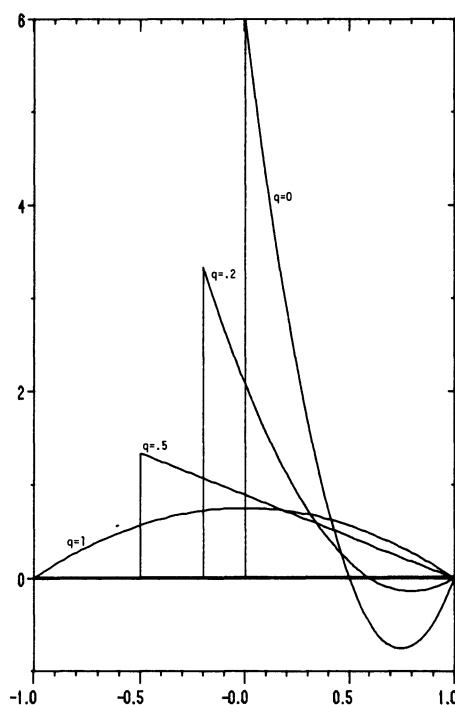


Figure 4. – Effets de bord : noyaux optimaux d'ordre 2 de support $[-q, 1]$ et continus à droite.

Cette solution se généralise aisément à un ordre plus élevé. Pour les premiers ordres on trouve :

$$r = 1 \quad 1 - x$$

$$r = 2 \quad (1 - x) [1 + 4q^3 + 3q^4 - 2x(1 - 3q^2 - 2q^3)]$$

Pour les ordres pairs on retrouve bien les noyaux symétriques de la section précédente en prenant $q = 1$. Pour les ordres impairs on arrive aux noyaux optimaux que l'on aurait obtenus en ne fixant que $K(1) = 0$.

La figure 4 donne pour quelques valeurs de q l'allure du noyau d'ordre 2. Le cas $q = 0$ correspondant à x situé à l'extrémité gauche du domaine des abscisses est particulièrement intéressant. Le noyau est alors $6(1 - x)(1 - 2x)$, sa variance est égale à $24/5$ et son deuxième moment vaut $-1/10$. A titre de comparaison le noyau d'EPANICHNIKOV simplement tronqué donne une variance 4 fois inférieure mais s'accompagne d'un biais très fort puisque son premier moment vaut $3/8$.

5. EQUIVALENCE DE LA REGRESSION POLYNOMIALE PONDEREE

Nous avons jusqu'à présent dégagé les noyaux répondant à une certaine optimalité pour un point intérieur comme pour un point de bord, ceci dans le cas d'une répartition régulière des abscisses. La méthode que nous allons présenter permet de retomber sur ces noyaux par une approche plus simple sur le plan des calculs comme sur le plan intuitif. De plus elle se prête à la généralisation au cas d'une répartition quelconque des abscisses.

Cette méthode que nous appellerons "régression polynômiale mobile" ou RPM consiste, pour l'estimation en x , à simplement ajuster un polynôme de degré m fixé sur les y_i dont les abscisses x_i tombent dans la fenêtre $[x - \delta, x + \delta]$. L'ajustement se fait au sens des moindres carrés usuels, avec éventuellement des pondérations v_i . La valeur en x du polynôme ajusté donne l'estimateur par RPM de $\mu_y(x)$. L'ajustement est "mobile" par le fait que, la fenêtre se déplaçant lorsque x se déplace, il est réadapté aux points qu'il trouve dans la nouvelle fenêtre.

Le principe de la méthode est bien connu pour le calcul de la tendance d'une série chronologique, contexte où elle est généralement dénommée "lissage par régression polynômiale" (voir par exemple KENDALL et STUART, 1966). Ce principe a été repris par CLEVELAND (1979) qui y introduisit la possibilité de pondérations dans les moindres carrés en ne s'intéressant toutefois pas à l'allure de ces pondérations et en mettant l'accent sur une procédure itérative robuste.

Les développements qui vont suivre conduisent à démontrer que la RPM produit en fait, de façon implicite, les noyaux optimaux des sections précédentes, que x soit intérieur ou qu'il soit sur un bord (voir proposition 2). Pour simplifier les écritures nous supposons que $x = 0$; pour revenir au cas général il suffit de lire $x_i - x$ en lieu et place de x_i . Reprenons l'estimateur linéaire général (2.1) dont le biais et la variance sont donnés en (2.6) et (2.7) et étendons la définition de la section 3 à cette forme plus générale.

Définition

On dit que l'estimateur linéaire $\hat{\mu}_y(x) = (\sum w_i y_i) / (\sum w_i)$, où x est fixé, a un biais d'ordre m ssi on a :

$$\sum w_i x_i^k = \begin{cases} 0 & \text{si } k = 1, \dots, m-1 \\ \beta_m \neq 0 & \text{si } k = m. \end{cases}$$

Cette définition se justifie par le fait que le biais est alors éliminé pour l'approximation de Taylor au degré $m-1$ de $\mu_y(x)$ au voisinage de x comme on le vérifie immédiatement en partant de l'expression (2.6).

Proposition 1

L'estimateur de $\mu_y(x)$ obtenu par RPM de degré $m-1$ a un biais d'ordre m quelles que soient les pondérations utilisées dans l'ajustement.

Démonstration

Nous prendrons $m=3$, la généralisation étant immédiate. Soit v_i des poids affectés aux points d'abscisses x_i et soit $a_0 + a_1 t + a_2 t^2$ le polynôme à ajuster autour de $x=0$. En posant

$$\mu_k = \frac{\sum v_i x_i^k}{\sum v_i}$$

les coefficients a_0, a_1 et a_2 sont donnés par les équations normales :

$$\begin{cases} a_0 + \mu_1 a_1 + \mu_2 a_2 = \sum v_i y_i / \sum v_i \\ \mu_1 a_0 + \mu_2 a_1 + \mu_3 a_2 = \sum v_i x_i y_i / \sum v_i \\ \mu_2 a_0 + \mu_3 a_1 + \mu_4 a_2 = \sum v_i x_i^2 y_i / \sum v_i \end{cases}$$

L'estimation au point x est donnée par a_0 , valeur du polynôme pour $t=0$. Avec un coefficient de proportionnalité c , indépendant des y_i , on a :

$$a_0 = c \begin{vmatrix} \sum v_i y_i & \mu_1 & \mu_2 \\ \sum v_i x_i y_i & \mu_2 & \mu_3 \\ \sum v_i x_i^2 y_i & \mu_3 & \mu_4 \end{vmatrix}$$

d'où nous pouvons déduire le poids w_i affecté à y_i pour l'estimateur linéaire $\sum w_i y_i / \sum w_i$ induit par la RPM de degré 2 et de pondérations v_i :

$$w_i = v_i \begin{vmatrix} 1 & \mu_1 & \mu_2 \\ x_i & \mu_2 & \mu_3 \\ x_i^2 & \mu_3 & \mu_4 \end{vmatrix}. \quad (5.19)$$

On voit immédiatement que les w_i vérifient quels que soient les v_i :

$$\sum w_i x_i = \sum w_i x_i^2 = 0. \quad (5.20)$$

L'intérêt de cette proposition est qu'elle reste valable sans aucune hypothèse de répartition des x_i et qu'elle laisse une liberté de choix des pondérations v_i que l'on pourra exploiter pour contrôler la variance. Nous nous contenterons de montrer qu'il existe un choix permettant de retomber sur les noyaux optimaux d'ordre m (ou éventuellement $m + 1$ si m est impair) des secteurs précédentes.

Proposition 2

Les assertions suivantes sont vraies à l'ordre d'approximation $O(1/n)$. Le noyau optimal (au sens des sections 3 et 4) d'ordre r est obtenu par RPM au degré $r - 1$ avec pondérations uniformes. Le noyau optimal du bord gauche (respectivement droit) continu à la borne droite (resp. gauche) est obtenu avec pondérations de la forme $1 - x$ (resp. $1 + x$). Pour un point intérieur le noyau optimal d'ordre $2k$ continu aux bornes est obtenu par RPM au degré $2k - 2$ avec pondérations $1 - x^2$ ou de façon équivalente par RPM au degré $2k - 1$ avec pondération $1 - x$ (ou encore $1 + x$).

Démonstration

Supposons pour simplifier que $\delta = 1$ (pour le cas général on substitue x_i/δ à x_i) et montrons que le choix $v_i = 1$ pour un point dans la fenêtre et $v_i = 0$ pour un point extérieur donne les noyaux optimaux sans contrainte de continuité aux bornes. Nous raisonnerons dans le cas d'un point de bord avec la fenêtre $[-q, 1]$, le cas d'un point intérieur équivalant à $q = 1$. En remplaçant v_i par 1 ou 0 dans (5.19), en transposant ensuite le déterminant et en généralisant à dimension m , on voit que le noyau implicite de la RPM est alors un polynôme de degré m proportionnel à :

$$P(x) = \begin{vmatrix} 1 & x & x^2 & \dots & x^m \\ \tilde{\rho}_1 & \tilde{\rho}_2 & \tilde{\rho}_3 & \dots & \tilde{\rho}_{m+1} \\ \cdot & & & & \\ \cdot & & & & \\ \cdot & & & & \\ \tilde{\rho}_m & \dots & \dots & \dots & \tilde{\rho}_{2m} \end{vmatrix} \quad (5.21)$$

où
$$\tilde{\rho}_k = \sum_{i \in J} x_i^k \quad \text{et} \quad J = \{i | x_i \in [-q, 1]\}$$

Comparons cette expression à (4.16) où ρ_k est l'intégrale de x^k sur $[-q, 1]$ c'est-à-dire, à une constante près, l'approximation en $O(1/n)$ de $\tilde{\rho}_k$. Dans (4.16) le vecteur $\tilde{\alpha}$ contient les coefficients du noyau polynômial de degré $r - 1$ optimal au sens défini précédemment. Le coefficient du terme de degré k est le cofacteur de l'élément de la $(k + 1)$ -ième colonne sur la première ligne et on constate qu'il s'agit bien du même coefficient dans (5.21) avec $m = r - 1$.

Considérons à présent v_i proportionnel à $1 - x_i$ dans la fenêtre et calculons le noyau ainsi obtenu implicitement par la RPM toujours pour un point du bord

gauche et à un coefficient de proportionnalité près; pour simplifier prenons $m = 2$. On a alors avec des coefficients c et $\tilde{c}$ indépendants de k :

$$\mu_k = c \sum v_i x_i^k = \tilde{c} (\tilde{\rho}_k - \tilde{\rho}_{k+1})$$

ce qui donne par (5.19) le noyau proportionnel à :

$$Q(x) = \begin{vmatrix} 1-x & x-x^2 & x^2-x^3 \\ \tilde{\rho}_1 - \tilde{\rho}_2 & \tilde{\rho}_2 - \tilde{\rho}_3 & \tilde{\rho}_3 - \tilde{\rho}_4 \\ \tilde{\rho}_2 - \tilde{\rho}_3 & \tilde{\rho}_3 - \tilde{\rho}_4 & \tilde{\rho}_4 - \tilde{\rho}_5 \end{vmatrix}$$

Mettons en évidence, par exemple, le coefficient du terme en x :

$$\begin{vmatrix} -1 & 1 & 0 \\ \tilde{\rho}_1 - \tilde{\rho}_2 & \tilde{\rho}_2 - \tilde{\rho}_3 & \tilde{\rho}_3 - \tilde{\rho}_4 \\ \tilde{\rho}_2 - \tilde{\rho}_3 & \tilde{\rho}_3 - \tilde{\rho}_4 & \tilde{\rho}_4 - \tilde{\rho}_5 \end{vmatrix} = \begin{vmatrix} -1 & 0 & 0 \\ \tilde{\rho}_1 - \tilde{\rho}_2 & \tilde{\rho}_1 - \tilde{\rho}_3 & \tilde{\rho}_3 - \tilde{\rho}_4 \\ \tilde{\rho}_2 - \tilde{\rho}_3 & \tilde{\rho}_2 - \tilde{\rho}_4 & \tilde{\rho}_4 - \tilde{\rho}_5 \end{vmatrix}$$

lequel est identique à α_1 dans (4.18). Ainsi par pondérations du type $1-x$ dans la RPM de degré 2 nous retrouvons le noyau continu à droite d'ordre 3 de la section précédente. Ceci se généralise sans peine à un ordre quelconque. Rappelons que pour un point intérieur, en raison de la symétrie, ceci ne peut valoir que pour un ordre pair. Dans ce cas les pondérations de la forme $1-x$ (ou $1+x$) avec une RPM de degré 3 donnent le noyau symétrique continu optimal d'ordre 4. On peut démontrer aisément que ceci peut être également obtenu par des pondérations de la forme $1-x^2$ avec une RPM de degré 2. On notera que l'approximation mentionnée dans la proposition 2 vient uniquement de l'approximation d'une somme de termes en x_i^k par une intégrale $\int x^k dx$.

6. GENERALISATIONS

Pour la recherche d'optimalité des noyaux il a été commode d'admettre une répartition régulière des abscisses x_i . Pour une répartition quelconque définissons la fonction :

$$G_n(t) = \frac{1}{n} \sum_{i=1}^n 1_{(-\infty, t]}(x_i)$$

Les expressions (2.6) et (2.7) du biais et de la variance en x , où les w_i sont issus d'un noyau K de support inclus dans $[-1, 1]$, s'écrivent alors :

$$\begin{aligned} \text{Biais} [\hat{\mu}_y(x)] &= \int_{x-\delta}^{x+\delta} K\left(\frac{t-x}{\delta}\right) [\mu_y(t) - \mu_y(x)] dG_n(t) \cdot \\ &\quad \cdot \left[\int_{x-\delta}^{x+\delta} K\left(\frac{t-x}{\delta}\right) dG_n(t) \right]^{-1} \end{aligned} \quad (6.22)$$

$$\text{Var} [\hat{\mu}_y(x)] = \frac{\sigma^2}{\delta} \int_{x-\delta}^{x+\delta} K^2\left(\frac{t-x}{\delta}\right) dG_n(t) \cdot \left[\int_{x-\delta}^{x+\delta} K\left(\frac{t-x}{\delta}\right) dG_n(t) \right]^{-2} \quad (6.23)$$

La solution optimale pour K va dépendre de G_n et ceci d'une façon complexe. Toutefois la RPM, dont on a montré qu'elle équivaut à prendre les noyaux optimaux dans le cas d'une répartition régulière, se prête à la généralisation.

Nous avons déjà vu, à la section précédente, que la proposition 1 relative au biais de la RPM reste vraie pour une répartition quelconque. En fait on montre même (LEJEUNE, 1984) que la RPM de degré $m - 1$ et de pondérations uniformes dans une fenêtre de largeur donnée, fournit l'estimateur de variance minimale dans la classe des estimateurs linéaires à biais d'ordre m . La RPM est donc une procédure qui, de par son caractère adaptatif, garantit l'optimalité au sens où nous l'entendons ici.

En pratique on observe que l'élimination du biais à un ordre d'approximation du deuxième degré suffit et on pourra se contenter d'une régression quadratique. Les pondérations uniformes peuvent induire des discontinuités fâcheuses de la courbe de régression estimée là où la concentration de points est faible. On préférera donc utiliser des pondérations selon une fonction de poids dérivable sur R afin de privilégier le lissage local de cette courbe, quitte à perdre un peu sur la variance (dans le cas d'une répartition régulière la variance est accrue ainsi de 25 %). Le choix de la fonction de poids :

$$v_i = \begin{cases} \left[1 - \left(\frac{x_i - x}{\delta} \right)^2 \right]^2 & \text{si } |x_i - x| < \delta \\ 0 & \text{sinon} \end{cases}$$

est particulièrement avantageux pour les calculs car il permet d'utiliser un algorithme de réactualisation rapide quand on incrémente le long d'une suite d'abscisses (voir LEJEUNE, 1984, ainsi que le programme en annexe).

Dans les figures 5 à 8 nous donnons quelques résultats de simulations avec des abscisses aléatoires de loi uniforme sur $[0, 1]$ et des erreurs gaussiennes. Les estimations obtenues sont très satisfaisantes car pour des tailles d'échantillons égales à 100, elles ne s'écartent que faiblement de la vraie courbe.

Pour ces estimations la largeur de fenêtre a été fixée entre 40 et 80 % de l'étendue totale du domaine des abscisses. Il est bien clair qu'un tel choix nécessite une certaine connaissance (ou exigence) a priori de l'ampleur d'oscillation de la courbe de régression. Ainsi pour l'estimation du spectre d'une série chronologique à partir de son périodogramme, une valeur de 15 à 20 % sera plus appropriée. Ces indications peuvent sembler artificielles mais on remarquera, et ceci est également essentiel pour justifier les noyaux d'ordre élevé, que la réduction du biais va de pair avec la réduction de la sensibilité de l'estimation vis-à-vis de la largeur de fenêtre. En effet en reprenant l'expression (2.8) du biais et en la dérivant par rapport à δ on obtient :

$$\int_{-1}^1 u K(u) \mu'_y(x + \delta u) du$$

ce qui, pour un noyau d'ordre r , s'annule comme le biais pour l'approximation locale de μ_y par un polynôme de degré $r - 1$. En pratique on observe effectivement que l'estimation résiste mieux à la variation de δ pour $r > 2$. Quant à la variance elle varie avec δ de façon indépendante de la forme du noyau. Nous ne nous sommes pas intéressé ici à la détermination du paramètre δ mais nous sommes conscient de la nécessité des investigations sur ce point et la rapidité de l'algorithme mentionné plus haut permet d'envisager, outre l'approche empirique par essais et erreurs, une approche du type validation croisée.

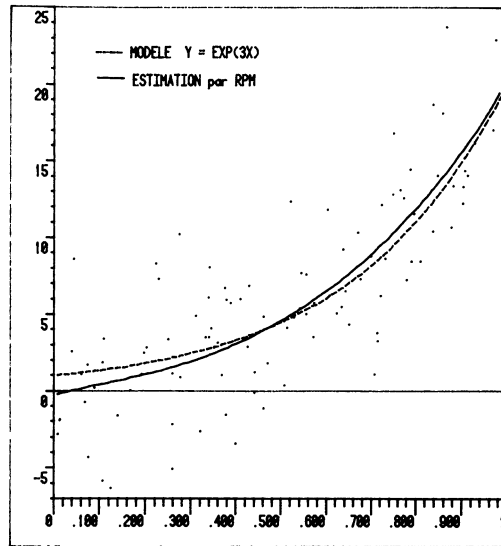


Figure 5. — Estimation ($\delta = 0.8$) à partir de 100 observations d'un modèle exponentiel (erreurs gaussiennes $\sigma = 4$) par la méthode de la régression polynômiale mobile de degré 2 avec pondérations en $(1 - u^2)^2$. Abcisses issues d'une loi uniforme sur $[0, 1]$.

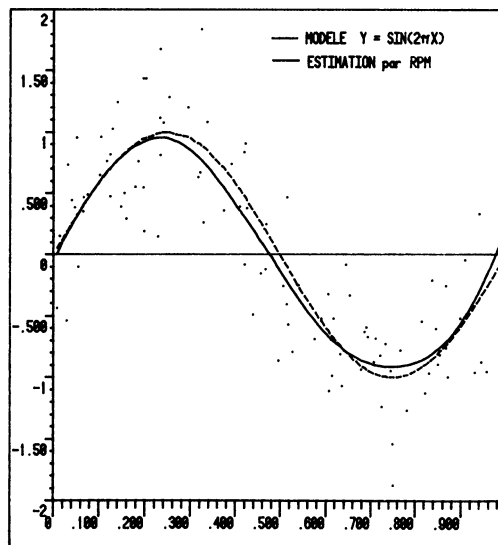


Figure 6. — Estimation ($\delta = 0.4$) à partir de 100 observations d'un modèle sinusoïdal (erreurs gaussienne $\sigma = 0.4$) par la méthode de la régression polynômiale mobile de degré 2 avec pondérations en $(1 - u^2)^2$. Abcisses issues d'une loi uniforme sur $[0, 1]$.

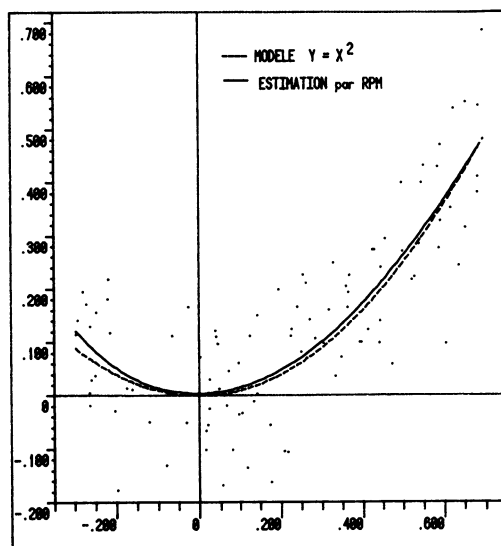


Figure 7. — Estimation ($\delta = 0.8$) à partir de 100 observations d'un modèle parabolique (erreurs gaussiennes $\sigma = 0.1$) par la méthode de la régression polynômiale mobile de degré 2 avec pondérations en $(1 - u^2)^2$. Abscisses issues d'une loi uniforme sur $[-.3, +.7]$.

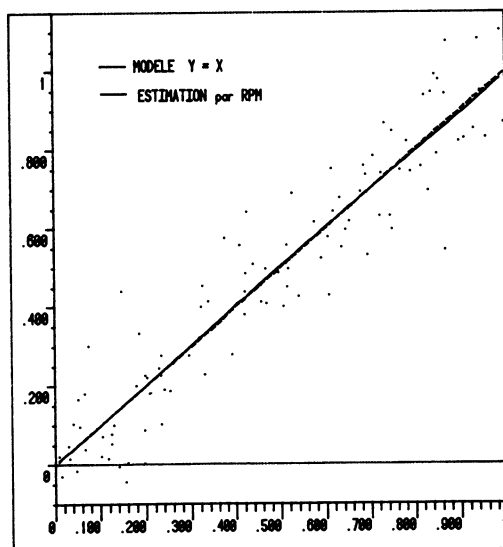


Figure 8. — Estimation ($\delta = 0.8$) à partir de 100 observations d'un modèle linéaire (erreurs gaussiennes $\sigma = 0.1$) par la méthode de la régression polynômiale mobile de degré 2 avec pondérations en $(1 - u^2)^2$. Abscisses issues d'une loi uniforme sur $[0, 1]$.

Finalement nous pouvons nous interroger sur l'incidence d'autres hypothèses posées initialement pour le modèle.

Levons en premier lieu l'hypothèse des x_i certains et supposons donc qu'ils soient des réalisations indépendantes d'une variable aléatoire de fonction de répartition G , revenant ainsi au modèle général de la régression. Le biais se calcule en prenant l'espérance mathématique de l'expression (6.22) où G_n est maintenant la fonction de répartition empirique reposant sur les x_i . Cette expression étant constituée d'un rapport de deux variables aléatoires on doit se contenter d'une approximation qui consiste simplement à remplacer G_n par G , laquelle a été justifiée par COLLOMB (1977) dans le cas d'un noyau positif. La proposition 1 de la section 5 implique ici que le biais conditionnel, les x_i étant donnés, est d'ordre m pour la RPM de degré $m - 1$. L'expression $[\mu_Y(t) - \mu_Y(x)]$ peut être remplacée par le reste de Taylor $R_m(t - x)$ dans (6.22) et le calcul de l'espérance mathématique ne retient finalement que ce reste. Si l'approximation de Taylor est bonne on peut obtenir une réduction substantielle du biais.

Quant aux hypothèses de l'indépendance des observations et de la constance de la variance σ^2 en tout x elles n'interviennent pas dans le calcul du biais. En conclusion on peut dire que la RPM privilégie la réduction du biais dans des situations variées. La variance de l'estimateur n'est pas nécessairement optimisée mais on peut espérer qu'elle reste dans des limites acceptables.

REFERENCES BIBLIOGRAPHIQUES

- W.S. CLEVELAND (1979). — Robust locally weighted regression and smoothing scatter plots, *JASA*, 74, 829-836.
- G. COLLOMB (1977). — Quelques propriétés de la méthode du noyau pour l'estimation non paramétrique de la régression en un point fixé, *C.R. Acad. Sci. Paris*, t. 285, Série A, 289-292.
- Y. DODGE, M. LEJEUNE (1982). — Some difficulties involving non parametric estimation of a density function, *Cahiers de Méthodes Quantitatives*, Tech. Report No 2 for FNRS, 1-24, Université de Neuchâtel.
- V.A. EPANECHNIKOV (1969). — Nonparametric estimation of a multidimensional probability density. *Theory Prob. Appl.*, 14, 153-158.
- T. GASSER, H.G. MULLER (1979). — Kernel estimation of regression functions, in *Smoothing Techniques for curve estimation*, ed. Gasser t. & Rosenblatt. Springer-Verlag Heidelberg, 23-68.
- T. GASSER, H.G. MULLER (1982). — Estimating regression functions and their derivatives by the kernel method, Mimeographed manuscript.
- M.G. KENDALL, A. STUART (1966). — *The advanced theory of statistics*, Vol. 3, § 46.5, Griffin.
- M. LEJEUNE (1984). — Optimization in non parametric regression. Compstat 84, *Proceedings in Computational statistics*, 412-426, Physica Verlag, Vienna.
- V. MAMMITZSCH (1983). — A note on kernel estimators fulfilling certain moment conditions, *Actes de la 44^e session de l'Inst. Int. de Stat. à Madrid*, Vol. 1, 30-33.
- R.A. TAPIA, J.R. THOMPSON (1980). — *Nonparametric density estimation*, Johns Hopkins University Press, Baltimore.

```

SUBROUTINE REGLISSE(X,Y,N,XT,YT,M,D)
C
C CALCUL DE L'ESTIMATION NON PARAMETRIQUE DE REGRESSION PAR REGRESSION
C POLYNOMIALE MOBILE DU SECOND DEGRE PONDEREE SELON (1-X**2)**2
C MICHEL LEJEUNE, UNIVERSITE DE LAUSANNE (1984)
C (X(I),Y(I),I=1,N) : nuage des observations
C (XT(J),J=1,M) : suite croissante d'abscisses où effectuer l'estimation
C (YT(J),J=1,M) : estimation des ordonnées
C D : demie largeur de fenêtre
C -----
C DIMENSION X(N),Y(N),XT(M),YT(M)
C DIMENSION S(8),SY(6),H(8)
C Tri du nuage en abscisses croissantes :
C CALL TRICLE(X,Y,N)
C Phase initiale
C P=-2./D**2
C Q=D**4
C X0=XT(1)
C DO 4 L=1,8
4 S(L)=0.
C DO 7 L=1,6
7 SY(L)=0.
C SY0=0.
C DO 1 I=1,N
C DI=X(I)-X0
C IF (DI.GE.-D)GOTO 2
1 CONTINUE
C CALL ARRETVIDE(X0)
2 IF(DI.GT.D) CALL ARRETVIDE(X0)
C N1=I
C DO 3 I=N1,N
C DI=X(I)-X0
C IF (DI.GT.D) GOTO 5
C CALL SOMME(S,SY0,SY,1.,Y(I),DI)
3 CONTINUE
C N2=N
C GOTO 6
5 N2=I-1
6 S0=N2-N1+1
C CALL FORMULE(S0,S,SY0,SY,YT(1),X0,P,Q)
C IF(M.EQ.1)RETURN
C phase courante
C DO 100 J=2,M
C X0=XT(J)
C H(1)=XT(J-1)-XT(J)
C DO 30 L=2,8
30 H(L)=H(L-1)*H(1)
C S(8)=S(8)+H(8)*S0+8.*(H(1)*S(7)+H(7)*S(1))+28.*(H(2)*S(6)+H(6)*
& S(2))+56.*(H(3)*S(5)+H(5)*S(3))+70.*H(4)*S(4)
C S(7)=S(7)+H(7)*S0+7.*(H(1)*S(6)+H(6)*S(1))+21.*(H(2)*S(5)+H(5)*
& S(2))+35.*(H(3)*S(4)+H(4)*S(3))
C S(6)=S(6)+H(6)*S0+6.*(H(1)*S(5)+H(5)*S(1))+15.*(H(2)*S(4)+H(4)*
& S(2))+20.*H(3)*S(3)
C S(5)=S(5)+H(5)*S0+5.*(H(1)*S(4)+H(4)*S(1))+10.*(H(2)*S(3)+H(3)*
& S(2))
C S(4)=S(4)+H(4)*S0+4.*(H(1)*S(3)+H(3)*S(1))+6.*H(2)*S(2)
C S(3)=S(3)+H(3)*S0+3.*(H(1)*S(2)+H(2)*S(1))
C S(2)=S(2)+H(2)*S0+2.*H(1)*S(1)
C S(1)=S(1)+H(1)*S0
C SY(6)=SY(6)+H(6)*SY0+6.*(H(1)*SY(5)+H(5)*SY(1))+15.*(H(2)*SY(4)+
& H(4)*SY(2))+20.*H(3)*SY(3)

```



```

      SY(5)=SY(5)+H(5)*SY0+5.*(H(1)*SY(4)+H(4)*SY(1))+10.*(H(2)*SY(3)+
&   H(3)*SY(2))
      SY(4)=SY(4)+H(4)*SY0+4.*(H(1)*SY(3)+H(3)*SY(1))+6.*(H(2)*SY(2)
      SY(3)=SY(3)+H(3)*SY0+3.*(H(1)*SY(2)+H(2)*SY(1))
      SY(2)=SY(2)+H(2)*SY0+2.*(H(1)*SY(1)
      SY(1)=SY(1)+H(1)*SY0
      DO 10 I=N1,N
      DI=X(I)-X0
      IF(DI.GE.-D) GOTO 20
      CALL SOMME(S,SY0,SY,-1.,Y(I),DI)
10    CONTINUE
      CALL ARRETVIDE(X0)
20    IF(DI.GT.D) CALL ARRETVIDE(X0)
      N1=I
      IF(N2.EQ.N) GOTO 60
      NN2=N2+1
      DO 40 I=NN2,N
      DI=X(I)-X0
      IF (DI.GT.D) GOTO 50
      CALL SOMME(S,SY0,SY,1.,Y(I),DI)
40    CONTINUE
      N2=N
      GOTO 60
50    N2=I-1
60    S0=N2-N1+1
      CALL FORMULE(S0,S,SY0,SY,YT(J),X0,P,Q)
100  CONTINUE
      RETURN
      END

```

```

C      SUBROUTINE ARRETVIDE(X0)
      PRINT *, ' LISSAGE IMPOSSIBLE EN X=',X0, '   FENETRE VIDE - STOP'
      STOP
      END

```

```

C      SUBROUTINE SOMME(S,SY0,SY,SIGNE,Y,DI)
      DIMENSION S(8),SY(6)
      DO 1 L=1,8
1     S(L)=S(L)+SIGNE*DI**L
      DO 2 L=1,6
2     SY(L)=SY(L)+SIGNE*DI**L*Y
      SY0=SY0+SIGNE*Y
      RETURN
      END

```

```

C      SUBROUTINE FORMULE(S0,S,SY0,SY,YT,X0,P,Q)
      DIMENSION T(4),S(8),SY(6)
      T0=Q*S(4)+P*S(2)+S0
      DO 1 L=1,4
1     T(L)=Q*S(L+4)+P*S(L+2)+S(L)
      TY0=Q*SY(4)+P*SY(2)+SY0
      TY1=Q*SY(5)+P*SY(3)+SY(1)
      TY2=Q*SY(6)+P*SY(4)+SY(2)
      A=T(2)*T(4)-T(3)*T(3)
      B=T(2)*T(3)-T(1)*T(4)
      C=T(1)*T(3)-T(2)*T(2)
      W=A*T0+B*T(1)+C*T(2)
      IF(W.LE.0.) CALL ARRETVIDE(X0)
      YT=(A*TY0+B*TY1+C*TY2)/W
      RETURN
      END

```

```

SUBROUTINE MINMAX(X,N,XMIN,XMAX)
DIMENSION X(N)
XMIN = X(1)
XMAX = X(1)
DO 45 I=1,N
IF( X(I) .LT. XMIN ) XMIN = X(I)
IF( X(I) .GT. XMAX ) XMAX = X(I)
45 CONTINUE
RETURN
END

```

```

C
SUBROUTINE TRICLE(X, XX, N)
C Tri du nuage (X(I),XX(I),I=1,N)
C selon les X(I) croissants.
C Algorithme QUICKSORT
C François Grize (Univ.Neuchâtel)
C Taille des piles :
PARAMETER M = 100
REAL X(N),XX(N)
INTEGER LEFT(M),RIGHT(M)
INTEGER S,L,R,I,J
REAL Y,Z
S=1
LEFT(1)=1
RIGHT(1)=N
1 L=LEFT(S)
R=RIGHT(S)
S=S-1
2 I=L
J=R
Y=X((L+R) / 2)
3 IF (X(I) .GE. Y) GOTO 4
I=I+1
GOTO 3
4 IF (Y .GE. X(J)) GOTO 5
J=J-1
GOTO 4
5 IF (I .LE. J) THEN
Z=X(I)
X(I)=X(J)
X(J)=Z
Z=XX(I)
XX(I)=XX(J)
XX(J)=Z
I=I+1
J=J-1
ENDIF
IF (I .LE. J) GOTO 3
IF (I .LT. R) THEN
S=S+1
LEFT(S)=I
RIGHT(S)=R
ENDIF
R=J
IF (L .LT. R) GOTO 2
IF (S .NE. 0) GOTO 1
RETURN
END
C

```