SEMI-RECURSIVE KERNEL ESTIMATION OF FUNCTIONS OF DENSITY FUNCTIONALS AND THEIR DERIVATIVES

A.V. Kitayeva and G.M. Koshkin

International Management Institute,
Tomsk Politecnic University, 30 *Lenin ave.,* 634050 *Tomsk, Russia*
e-mail: olz@tomsknet.ru
Department of Applied Mathematics and Cybernetics,
Tomsk State University, 36 *Lenin ave.,* 634050 *Tomsk, Russia*
e-mail: kgm@mail.tsu.ru

Abstract: A class of semi-recursive kernel type estimates of functions depending on multivariate density functionals and their derivatives is considered. The piecewise smoothed approximations of these estimates are proposed. The convergence with probability one of the estimates is proved. The main parts of the asymptotic mean square errors of the estimates are found. The examples of estimation of the production function, the marginal productivity and the marginal rate of technical substitution of inputs are given.

Keywords: Kernel estimate, recursive estimation, mean square convergence, almost surely convergence.

1. INTRODUCTION

Solution of many nonparametric statistical problems (such as identification, classification, filtering, prediction, etc.) is based on estimation of certain probabilistic characteristics of the following type expressions:

$$J(x) =$$

$$= H\left(\{a_i(x)\},\ \{a_i^{(1j)}(x)\},\ i=\overline{1,s},\ j=\overline{1,m}\right) =$$

$$= H\left(a(x),\ a^{(1j)}(x)\right). \qquad (1)$$

Here $x \in \mathsf{R}^m$, $H(t) : \mathsf{R}^{(m+1)s} \to \mathsf{R}^1$ is a given function,

$$a^{(0j)}(x) = a(x) = (a_1(x), \dots, a_s(x)),$$

$$a^{(1j)}(x) = \left(a_1^{(1j)}(x), \dots, a_s^{(1j)}(x)\right),$$

$$a_i(x) = \int g_i(y) f(x,y)dy, \quad i=\overline{1,s},$$

$$a_i^{(1j)}(x) = \frac{\partial a_i(x)}{\partial x_j}, \quad i=\overline{1,s}, \quad j=\overline{1,m},$$

where $g_1, \dots, g_s$ are known Borel functions, $\int \equiv \int_{\mathsf{R}^1}$, $f(x,y)$ is an unknown probability density function (p.d.f.) for the observed random vector $Z=(X,Y) \in \mathsf{R}^{m+1}$.

If $g_i(y) \equiv 1$, then

$$a_i(x) = \int f(x,y)dy = p(x),$$

where $p(x)$ is the marginal probability density of the random variable X, and $f(y|\,x) = \dfrac{f(x,y)}{p(x)}$ is a conditional probability density.

Remark 1. Note that in (1) some variables of function $H(\cdot)$ may be absent (for example all derivatives).

Here are some well known examples of such kind of functions:

– the regression line

$$r(x) = \int y f(y|x) dy,$$

$$(H(a_1, a_2) = \frac{a_1}{a_2}, \quad g_1(y) = y, \quad g_2(y) = 1);$$

– the conditional initial moments

$$\mu_m(x) = \int y^m f(y|x) dy,$$

$$(H(a_1, a_2) = \frac{a_1}{a_2}, \ g_1(y) = y^m, \ m \geq 1, \ g_2(y) = 1);$$

– the conditional variance

$$D(x) = \mu_2(x) - r^2(x),$$

$$(H(a_1, a_2, a_3) = a_1/a_3 - (a_2/a_3)^2, \quad g_1(y) = y^2,$$

$$g_2(y) = y, \quad g_3(y) = 1);$$

– the conditional standard deviation

$$\sigma(x) = \sqrt{D(x)},$$

$$(H(a_1, a_2, a_3) = \sqrt{a_1/a_3 - (a_2/a_3)^2});$$

– the conditional central moments

$$V_m(x) = \int (y - r(x))^m f(y|x) dy,$$

$$(g_1(y) = y, \quad g_2(y) = y^2, \ldots,$$

$$g_m(y) = y^m, \quad g_{m+1}(y) = 1);$$

– the conditional coefficient of skewness

$$\beta_1(x) = \frac{\mathsf{E}((Y - r(x))|x)^3}{[\mathsf{D}(Y|x)]^{3/2}},$$

$$H(a_1, a_2, a_3, a_4) = \frac{b_4 - 3b_3 b_2 + 2b_2^3}{(b_3 - b_2^2)^{3/2}},$$

$$b_i = \frac{a_i}{a_1}, \ q_i(y) = y^{i-1}, \ i = \overline{1, 4};$$

– the sensitivity functions. For example

$$T_j(x) = \frac{\partial r(x)}{\partial x_j}; \quad H(a_1, a_2, a_1^{(1j)}, a_2^{(1j)}) =$$

$$= \frac{a_1^{(1j)}}{a_2} - \frac{a_1 a_2^{(1j)}}{a_2^2} = b_1^{(1j)}, \ g_1(y) = y, \ g_2(y) = 1.$$

2. PROBLEM STATEMENT

Take the following expression as an estimate of the functional $a(x) = a^{(0j)}(x)$ $(r = 0)$ and its derivatives $a^{(1j)}(x)$ $(r = 1)$ at a point x:

$$a_n^{(rj)}(x) = \frac{1}{n} \sum_{i=1}^{n} \frac{g(Y_i)}{h_i^{m+r}} \mathbf{K}^{(rj)} \left(\frac{x - X_i}{h_i} \right). \qquad (2)$$

Here $Z_i = (X_i, Y_i)$, $i = \overline{1, n}$, is the $(m+1)$-dimensional random sample from p.d.f. $f(x, y)$, (h_i) is a sequence of positive bandwidths tending to 0 as $i \to \infty$, $\mathbf{K}^{(0j)}(u) = \mathbf{K}(u) = \prod_{i=1}^{m} K(u_i)$ is a kernel m-dimensional multiplicative function which does not need to possess the characteristic properties of p.d.f.,

$$\mathbf{K}^{(1j)}(u) = \frac{\partial \mathbf{K}(u)}{\partial u_j} =$$

$$= K(u_1) \cdots K(u_{j-1}) K^{(1)}(u_j) K(u_{j+1}) \cdots K(u_m),$$

$$K^{(1)}(u_j) = \frac{dK(u_j)}{du_j},$$

$$a_n^{(rj)}(x) = \left(a_{1n}^{(rj)}(x), \ldots, a_{sn}^{(rj)}(x) \right),$$

$$g(y) = (g_1(y), \ldots, g_s(y)).$$

Note that (2) can be computed recursively by

$$a_n^{(rj)}(x) = a_{n-1}^{(rj)}(x) -$$

$$-\frac{1}{n} \left[a_{n-1}^{(rj)}(x) - \frac{g(Y_n)}{h_n^{m+r}} \mathbf{K}^{(rj)} \left(\frac{x - X_n}{h_n} \right) \right]. \qquad (3)$$

This property is particularly useful in large sample size since (3) can be easily updated with each additional observation. The recursive kernel estimate of $p(x)$ $(m = 1, \quad s = 1, \quad g(y) = 1, \quad H(a_1) = a_1)$ was introduced by Wolverton and Wagner (1969) and apparently independently by Yamato (1971), and has been thoroughly examined in (Wegman and Davies, 1979). Semi-recursive kernel type estimates of conditional functionals

$$b(x) = (b_1(x), \ldots, b_{s-1}(x))$$

$$b_i(x) = a_i(x)/p(x) = \int g_i(y) f(y|x) dy$$

at a point x are designed as

$$b_n(x) = \frac{\sum_{i=1}^{n} \frac{g(Y_i)}{h_i^m} \mathbf{K}\left(\frac{x - X_i}{h_i} \right)}{\sum_{i=1}^{n} \frac{1}{h_i^m} \mathbf{K}\left(\frac{x - X_i}{h_i} \right)} =$$

$$= \frac{a_n(x)}{p_n(x)} = \frac{a_n^{(0j)}(x)}{a_{sn}^{(0j)}(x)}, \quad g_s(x) = 1. \tag{4}$$

The substitution estimates are often used for the estimation of ratios. The possible unboundedness of the ratio estimates at some points (see (Cramér, 1975) for details) creates a difficulty in the job. Such estimates are called semi-recursive because they can be updated sequentially by adding extra terms to both the numerator and denominator when new observations became available. If $g_1(y) = y$ $(s = 2)$ we obtain semi-recursive kernel type estimates of the regression line (see (Ahmad and Lin, 1976; Buldakov and Koshkin, 1977; Devroye and Wagner, 1980)). Weak and strong universal consistency of such estimates was investigated in (Krzyźak, and Pawlak, 1984; Greblicki, and Pawlak, 1987; Krzyźak, 1992; Györfi, *et al.*, 1998; Walk, 2001). For estimation of (1) we are going to use the following statistic (the substitution estimate)

$$J_n(x) = H\left(\left\{a_n^{(rj)}(x)\right\}, \quad j = \overline{1,m}, \quad r = 0,1\right). \tag{5}$$

But the studying of the MSE for $J_n(x)$ has some difficulties due to the possible instability (for example, the denominator in (4) may be close to zero), and the theorems for MSE making use of the dominant sequences can not be applied (Cramér, 1975; Koshkin, 1999). The problem can be resolved by using a piecewise smooth approximation. Therefore, similar to (Penskaya, 1990; Koshkin, 1999) we use the estimate

$$\widetilde{J}_{n,\nu}(x) = \frac{J_n(x)}{(1 + \delta_{n,\nu}|J_n(x)|^{\tau})^{\rho}}, \tag{6}$$

where $\tau > 0$, $\rho > 0$, $\rho\tau \geq 1$, $(\delta_n) \downarrow 0$ as $n \to \infty$.

3. MEAN SQUARE ERRORS

Denote by $\sup\limits_x = \sup\limits_{x \in \mathsf{R}^m}$, $T_j = \int u^j K(u) du$, $j = 1, 2, \ldots$.

Definition 1. A function $H(z) : \mathsf{R}^s \to \mathsf{R}^1$ belongs to the class $\mathcal{N}_\nu(t)$ $(H(z) \in \mathcal{N}_\nu(t))$ if it is continuously differentiable up to the order ν at the point $t \in \mathsf{R}^s$. A function $H(z) \in \mathcal{N}_\nu(\mathsf{R})$ if it is continuously differentiable up to the order ν for any $z \in \mathsf{R}^s$.

Definition 2. A Borel function $K(u) \in \mathcal{A}^{(r)}$, $r = 0, 1$ $(\mathcal{A}^{(0)} = \mathcal{A})$ if $\int |K^{(r)}(u)| du < \infty$, and $\int K(u) du = 1$.

Definition 3. A Borel function $K(u)$ belongs to the class $\mathcal{A}_\nu^{(r)}$ $(\mathcal{A}_\nu^{(0)} = A_\nu)$ if $K(u) \in \mathcal{A}^{(r)}$, $\int |u^\nu K(u)| du < \infty$, $T_j = 0$, $j = 1, \ldots, \nu - 1$, $T_\nu \neq 0$, and $K(u) = K(-u)$.

Definition 4. A sequence (h_n) belongs to the class $\mathcal{H}(m + r + q)$ if

$$(h_n + 1/(nh_n^{m+r+q})) \downarrow 0,$$

and

$$\frac{1}{n}\sum_{i=1}^{n} h_i^\lambda = S_\lambda h_n^\lambda + o(h_n^\lambda),$$

where λ is a real number, S_λ is some constant independent on n.

Definition 5. Let t_n, $X_1, \ldots, X_n$ are vectors, and $t_n = t_n(X_1, \ldots, X_n)$. A sequence of functions $\{H(t_n)\}$ belongs to the class $\mathcal{M}(\gamma)$ if for any possible values $X_1, \ldots, X_n$ the sequence $\{|H(t_n)|\}$ is dominated by a sequence of numbers $(C_0 d_n^\gamma)$, $(d_n) \uparrow \infty$ as $n \to \infty$, $0 \leq \gamma < \infty$, C_0 is a constant.

Put

$$A = A(x) = \left(\left\{a^{(rj)}(x)\right\}, r \in [0,1], j = \overline{1,m}\right), \tag{7}$$

$$H_{tjr} = \partial H(A)/\partial a_t^{(rj)};$$

$$H\left(\left\{a_n^{(rj)}(x)\right\}, \quad r \in [0,1] \quad j = \overline{1,m}\right) = H(A_n);$$

$$a_{t,p}(x) = \int g_t(y) g_p(y) f(x,y)\, dy;$$

$$a_{t,p}^{1+}(x) = \int |g_t(y) g_p(y)| f(x,y)\, dy, \quad t,p = \overline{1,s};$$

$$a^{s+}(x) = \int |g^s(y)| f(x,y) dy;$$

for $r, q = 0, 1$

$$L^{(r,q)} = \int K^{(r)}(u) K^{(q)}(u)\, du;$$

$$\mathcal{B}_{t,p}^{(r,q)} = L^{(r,q)} \left(L^{(0,0)}\right)^{m-1} a_{t,p}(x);$$

$$\omega_{i\nu}^{(rj)}(x) = \frac{T_\nu}{\nu!} \sum_{l=1}^{m} \frac{\partial^\nu a_i^{(rj)}(x)}{\partial x_l^\nu}.$$

Remark 2. Note that in (7) the notation $r \in [0,1]$ means $r = 0$ or $r = 1$, or $r = 0, 1$. For instance for the last example in Introduction $r = 0, 1$ and for the rest examples $r = 0$.

Let the set

$$Q = \begin{cases} \{0\} & \text{if } \forall j \;\; r = 0; \\ \{1\} & \text{if } \forall j \;\; r = 1; \\ \{0,1\} & \text{if } \exists j \;\; r = 0 \bigwedge r = 1 \end{cases}$$

in formula (7).

Theorem 1 (MSE of the estimate $J_n(x)$).
Let $t, p = \overline{1, s}$, $j = \overline{1, m}$, $r \in Q$,
1) functions $a_{t,p}(z) \in \mathcal{N}_0(\mathsf{R})$;
2) $\sup_x a_{t,p}^{1+}(x) < \infty$, $\sup_x a_t^{1+}(x) < \infty$, $\sup_x a_t^{4+}(x) < \infty$;
3) the kernel function $K(u) \in \mathcal{A}_\nu^{(\max(r))}$, $\sup_x \left|K^{(r)}(x)\right| < \infty$, $K^{(r)}(z) \in \mathcal{N}_0(\mathsf{R})$ if the integer r takes both the value 0 and the value 1 (i. e. $Q = \{0, 1\}$), $\lim_{|u| \to \infty} K(u) = 0$ if $1 \in Q$;
4) $a_t^{(rj)}(z) \in \mathcal{N}_\nu(\mathsf{R})$, $\sup_x |a_t^{(rj)}(x)| < \infty$,
$\sup_x \left| \dfrac{\partial^\nu a_t^{(rj)}(x)}{\partial x_l \partial x_t \ldots \partial x_q} \right| < \infty$, $l, t, \ldots, q = \overline{1, m}$;
5) the sequence $(h_n) \in \mathcal{H}(m + 2\max(r))$.
Moreover, if
6) $H(z) \in \mathcal{N}_2(A)$;
7) $\{H(A_n)\} \in \mathcal{M}(\gamma)$, $0 \le \gamma \le 1/4$,
then MSE $u^2(J_n(x))$ of the estimate (1) can be written in the form

$$\mathsf{u}^2(J_n(x)) = \sum_{t,p=1}^{s} \sum_{j,k=1}^{m} \sum_{r,q \in Q} H_{tjr} H_{pkq} \times$$

$$\times \left[S_{-(m+2\max(r,q))} \frac{\mathcal{B}_{t,p}^{(r,q)}}{n h_n^{m+r+q}} + \right.$$

$$\left. + S_\nu^2\, \omega_{t\nu}^{(rj)}(x) \omega_{p\,\nu}^{(qk)}(x) h_n^{2\nu} \right] +$$

$$+ O\left(\left[\frac{1}{n h_n^{m+2\max(r)}} + h_n^{2\nu} \right]^{\frac{3}{2}} \right).$$

As above mentioned restriction 7) of Theorem 1 is the most problematical and we don't need one when piecewise smooth approximation (6) is used.

Theorem 2 (MSE of the piecewise smooth approximation estimate $\widetilde{J}_{n,\nu}(x)$). Suppose that conditions 1)–6) of Theorem 1 hold and restriction 7) is replaced by
7*) $J(x) = H(A(x)) \neq 0$ or $\tau \ge 4$,
τ is a positive integer. Then as $n \to \infty$

$$\mathsf{u}^2(\widetilde{J}_n(x)) \sim \mathsf{u}^2(J_n(x)).$$

4. ALMOST SURELY CONVERGENCE

Theorem 3. Suppose the conditions of Theorem 1 (or Theorem 2) hold, and, moreover,

$$\int |K^{(r)}(u)|\, du < \infty,$$

$$\sum_{n=1}^{\infty} \left(n^{-2} h_n^{-2(m+2\max(r))} + h_n^{4\nu} \right) < \infty.$$

Then the sequence of estimates $\{J_n(x)\}$ (or $\{\widetilde{J}_n(x)\}$) as $n \to \infty$ converges almost surely to $J(x)$.

5. ESTIMATION OF PRODUCTION FUNCTION AND ITS CHARACTERISTICS

Example 1. Let $r(x)$, $x = (x_1, x_2) \in \mathsf{R}^2$ be a regression model of production function, $a(x) = (a_1(x), a_2(x))$, $a_1(x) = \int y f(x,y) dy$, $a_2(x) = \int f(x,y) dy = p(x)$. Here $x_1 > 0$ is a capital input, $x_2 > 0$ is a labor input, $y > 0$ is a product, and $f(x,y) > 0$ only if $x_1 > 0$, $x_2 > 0$, $y > 0$. Then

$$J_n(x) = r_n(x) = \frac{\sum_{i=1}^{n} \frac{Y_i}{h_i^2} \mathbf{K}\left(\frac{x - X_i}{h_i}\right)}{\sum_{i=1}^{n} \frac{1}{h_i^2} \mathbf{K}\left(\frac{x - X_i}{h_i}\right)} =$$

$$= \frac{a_{1n}^{(0j)}(x)}{a_{2n}^{(0j)}(x)} = \frac{a_{1n}(x)}{p_n(x)}. \qquad (8)$$

Let $\mathbf{K}(u) = K(u_1)K(u_2)$, $K(u) \in \mathcal{A}_\nu$, and $(h_n) \in \mathcal{H}(m)$. To find the MSE of the estimate $r_n(x)$, we use Theorem 1. In view of 1)–4) conditions of the theorem we have: functions $a_i(x)$, $i = 1, 2$, and their derivatives are continuously differentiable up to the order ν for any $z \in \mathsf{R}^2$, and the function $\int y^4 f(x,y) dy$ is bounded on R^2. Also $\sup_{u \in \mathsf{R}^1} |K(u)| < \infty$. If $p(x) > 0$, then condition 6) is fulfilled obviously. We can not find the dominant sequence (d_n) (condition 7)) of Theorem 1, since the denominator in (8) may be equal to zero. Therefore it seems impossible to determine the MSE of the estimation $r_n(x)$, $\nu > 2$ (Nadaraya, 1964; Nadaraya, 1965; Nadaraya, 1983; Collomb, 1976). But it is shown in these papers that we can find the dominant sequence with $\gamma = 0$ under $\nu = 2$ according to Definition 5 if, for example, $K(u) \ge 0$, and $Y < \infty$. In this case as $n \to \infty$

$$\mathsf{u}^2(r_n(x)) = \sum_{i,p=1}^{2} H_i H_p \left(S_{-2} \frac{C B_{i,p}}{n h_n^2} + \right.$$

$$\left. + S_2^2\, \omega_{i2}(x) \omega_{p\,2}(x) h_n^4 \right) + O\left(\left[\frac{1}{n h_n^2} + h_n^4 \right]^{3/2} \right),$$

where

$$H_1 = \frac{1}{p(x)},\ H_2 = -\frac{r(x)}{p(x)};\ B_{1,1} = \int y^2 f(x,y) dy,$$

$$B_{1,2} = B_{2,1} = \int y f(x,y) dy, \quad B_{2,2} = p(x);$$

$$\omega_{12}(x) = \frac{T_2}{2}\left(\frac{\partial^2 a_1(x)}{\partial x_1^2} + \frac{\partial^2 a_1(x)}{\partial x_2^2}\right),$$

$$\omega_{22}(x) = \frac{T_2}{2}\left(\frac{\partial^2 p(x)}{\partial x_1^2} + \frac{\partial^2 p(x)}{\partial x_2^2}\right);$$

$$C = \int K^2(u) du.$$

For $\nu > 2$ we can use the piecewise smooth approximation $\widetilde{r}_n(x)$:

$$\widetilde{r}_n(x) = \frac{r_n(x)}{(1 + \delta_{n,\nu}| r_n(x)|^{\tau})^{\rho}}, \tag{9}$$

where $\tau > 0$, $\rho > 0$, $\rho\tau \geq 1$, $\delta_{n,\nu} = O\left(h_n^{2\nu} + 1/(nh_n^2)\right)$, $(\delta_{n,\nu}) \downarrow 0$ as $n \to \infty$. In view of condition 7) of Theorem 2 it is enough to take even $\tau \geq 4$, and as $n \to \infty$

$$\mathsf{u}^2(\widetilde{r}_n(x)) = \sum_{i,p=1}^{2} H_i H_p \left(S_{-2}\frac{CB_{i,p}}{nh_n^2} + S_\nu^2\, \omega_{i\nu}(x)\omega_{p\nu}(x) h_n^{2\nu}\right) + O\left(\left[\frac{1}{nh_n^2} + h_n^{2\nu}\right]^{3/2}\right),$$

where

$$\omega_{1\nu}(x) = \frac{T_\nu}{\nu!}\left(\frac{\partial^\nu a_1(x)}{\partial x_1^\nu} + \frac{\partial^\nu a_1(x)}{\partial x_2^\nu}\right),$$

$$\omega_{2\nu}(x) = \frac{T_\nu}{\nu!}\left(\frac{\partial^\nu p(x)}{\partial x_1^\nu} + \frac{\partial^\nu p(x)}{\partial x_2^\nu}\right).$$

Example 2. In the case of the marginal productivity function $T_1(x) = \dfrac{\partial r(x)}{\partial x_1}$ a dominant sequence finding difficulties force us to use the piecewise smooth approximation $\widetilde{T}_{1n}(x)$ at once:

$$\widetilde{T}_{1n}(x) = \frac{T_{1n}(x)}{(1 + \delta_n| T_{1n}(x)|^{\tau})^{\rho}},$$

where

$$T_{1n}(x) = \frac{1}{h_i}\left[\frac{\sum_{i=1}^{n} Y_i \mathbf{K}^{(11)}\left(\frac{x - X_i}{h_i}\right)}{\sum_{i=1}^{n} \mathbf{K}\left(\frac{x - X_i}{h_i}\right)} - \frac{\sum_{i=1}^{n} Y_i \mathbf{K}\left(\frac{x - X_i}{h_i}\right)\sum_{i=1}^{n} \mathbf{K}^{(11)}\left(\frac{x - X_i}{h_i}\right)}{\left[\sum_{i=1}^{n} \mathbf{K}\left(\frac{x - X_i}{h_i}\right)\right]^2}\right], \tag{10}$$

$\mathbf{K}^{(11)}(u) = K^{(1)}(u_1)K(u_2)$. The kernel $K(u)$ has to satisfy the complementary conditions (added to Example 1 restrictions): $\sup_{u \in \mathsf{R}^1} | K^{(1)}(u)| < \infty$, $\lim_{|u| \to \infty} K(u) = 0$, $K^{(\alpha)}(z) \in \mathcal{N}_0(\mathsf{R})$, $\alpha = 1, 2$; furthermore, the sequence (h_n) has to satisfy the condition $\lim_{n \to \infty} \left(h_n + (nh_n^3)^{-1}\right) = 0$. To use Theorem 2 result and obtain $\mathsf{u}^2\left(\widetilde{T}_{1n}(x)\right)$ functions $a_1(x)$, $a_2(x)$ and their derivatives up to the order $(\nu + 1)$ need to be continuous and bounded on R^2.

Example 3. Let $T_j(x) = \partial r(x)/\partial x_j$ and

$$MRTS_{12,n}(x) = T_{1n}(x)/T_{2n}(x)$$

be the estimate of the marginal rate of technical substitution of an input x_2 with an input x_1, where the denominator $T_{2n}(x)$ is given by (10), where $\mathbf{K}^{(11)}(u)$ is replaced by

$$\mathbf{K}^{(12)}(u) = K(u_1)K^{(1)}(u_2).$$

The piecewise smooth approximation of the estimate $MRTS_{12,n}(x)$ can be written easily. In view of condition 6) of Theorem 1 the inequality $r(x) \neq \dfrac{\partial a_1(x)}{\partial x_2} \Big/ \dfrac{\partial p(x)}{\partial x_2}$ have to hold in addition to Example 2 restrictions.

REFERENCES

Ahmad J.A., Lin P.E. (1976) Nonparametric sequential estimation of a multiple regression function. *Bull. Math. Statist.,* 17(1–2), 63–75.

Buldakov V.M.., Koshkin G.M. (1977). On the recursive estimates of a probability and a regression line. *Problems Inform. Trans.,* 13, 41–48.

Collomb G. (1976). *Estimation non parametrique de la regression par la methode du noyau.* Thèse Docteur Ingénieur. Toulouse: Univ. Paul-Sabatier.

Cramér H. (1948) *Mathematical Methods of Statistics.* Princeton.

Devroye L. and Wagner T.J. (1980). On the L_1 convergence of kernel estimates of regression functions with applications in discrimination. *Z. Wahrsch. Verw. Gebiete,* 51, 15–25.

Grebiicki W. and Pawlak M. (1987). Necessary and sufficient consistency conditions for recursive kernel regression estimate. *J. Multivariate Anal.*, 23, 67–76.

Györfi L., Kohler M. and Walk H. (1998). Weak and strong universal consistency of semi-recursive kernel and partitioning regression estimates. *Statist. Decisions,* 16, 1–18.

Koshkin G.M. (1999). Deviation moments of the substitution estimate and its piecewise smooth approximations. *Siberian Math. J.,* 40(3), 515–527.

Krzyżak A. (1992). Global convergence of the recursive kernel estimates with applications in classification and nonlinear system estimation. *IEEE Trans. Inform. Theory,* IT–38, 1323–1338.

Krzyżak A. and Pawlak M. (1984). Almost everywhere convergence of a recursive regression function estimate and classification. *IEEE Trans. Inform. Theory,* IT–30, 91–93.

Nadaraya E.A. (1964). On non-parametric regression estimator. *Theory Probab. Appl.,* 19(1), 147–149.

Nadaraya E.A. (1965). On non-parametric estimators of a regression curve. *Transactions of Comp. Center of AN of GSSR.* Tbilisi, Metsniereba, 242(5:1), 56–68.

Nadaraya E.A. (1983). *Non-parametric estimation of a probability density and a regression curve.* Tbilisi Univ. Publishing House, Tbilisi.

Penskaya M.Ja. (1990). On Stable Estimation of Function of Parameter. In: *Statistical Methods of Estimation and Testing of Hypothesis.* 44-55. Perm University Publishing House, Perm (in Russian).

Walk H. (2001). Strong universal pointwise consistency of recursive kernel regression estimates. *Ann. Inst. Statist. Math.*, 53(4), 691–707.

Wegman E.J. and Davies H.I. (1979). Remarks on some recursive estimates of a probability density function. *Ann. Statist.,* 7(2), 316-327.

Wolverton C.T., Wagner T.J. (1969). Recursive estimates of probability densities. *IEEE Trans. Syst. Sci. and Cybernet.*, SSC–5 (3), 246–247.

Yamato H. (1971). Sequential estimation of a continuous probability density function and mode. *Bulletin of Mathematical Statistics,* 14, 1–12.