PROBABILITY
AND
MATHEMATICAL STATISTICS

Vol. 12, Fasc. 1 (1991), pp. 1–24

ASYMPTOTIC NONPARAMETRIC SPLINE DENSITY ESTIMATION*

BY

Z. CIESIELSKI (SOPOT)

Abstract. In [5] we have announced a linear spline method for nonparametric density and distribution estimation on the real line. In this paper, asymptotic properties of a large family of such estimators are discussed. It is interesting that the results do not require the existence of derivatives of the density in question. For the asymptotic results on distribution functions, knowledge of the behavior of the second modulus of smoothness in the L^∞-norm is sufficient, and in the case of density estimation — knowledge of the behavior of the second modulus of smoothness in the L^1-norm and of the tail function is needed. The method of estimation is a kernel method which is not of convolution type. In the case of densities and of the L^1-norm it is as good as the optimal kernel methods (of convolution type) in an essentially larger class of density functions. Moreover, at the same time we get for free the estimator for the distribution function corresponding to the density in question. At the end we have derived, for a given sample, an explicit function of the window parameter. It is called a *window function* and it makes possible in each case to determine the size of the optimal window parameter. In obtaining the results the techniques of approximation theory, in particular by splines, are used following the same guidelines as presented in [7], [6] and [8].

1. Introduction. Density estimation is very much related to approximation of L^1-functions on the real line by positive linear operators. The nonparametric *kernel estimators* for densities correspond to approximation by operator given by kernels (cf. [15], [9], [11]). For densities with support in [0, 1] the asymptotic properties of the kernel estimators constructed by means of the *Bernstein-like* polynomial operators are discussed in [7], [6]. An asymptotic nonparametric kernel density estimator for densities supported on the d-dimensional cube $[0, 1]^d$ is discussed in [8], where the kernel is the Dirichlet kernel corresponding to the multidimensional Haar orthogonal system. Since in the polynomial case we deal with degenerate B-splines, i.e., the basic Bernstein polynomials, and in the Haar case with B-splines of the lowest order, i.e., of order 1, it was natural to ask for kernel estimators constructed by means of more smooth B-splines. Our method corresponds to a kernel, i.e., a kernel

* While this paper was in print its extension to several variables was presented by the author in the *Multivariate Approximation and Interpolation*, ISNM 94 (1990), Birkhäuser-Verlag, Basel, pp. 25–53.

function constructed by means of B-splines with equally spaced knots and it relies very much on the approximation theorem established in a particular but important case in [5]. In this note we present theorems on asymptotic order of approximation of the theoretical density (distribution) by the spline estimators and we discuss, given a simple finite sample, a method for *optimal choice* of the *window width* parameter.

2. The splines with equally spaced knots. In what follows we consider splines of order $r \geqslant 2$, i.e., of degree $r-1$, of maximal smothness and corresponding to the uniform mesh

$$\Pi_{h,\theta} = \{t_i = h(i+\theta), i \in Z\} \tag{2.1}$$

with step $h > 0$ and the real parameter θ which may depend on r, where Z is the set of all integers. It is assumed that all the knots are simple. For the properties listed below we refer to [13], [2] and [14].

All the B-splines corresponding to the mesh (2.1) can be defined by means of the *cardinal* B-spline of order r

$$N^{(r)}(x) = r[0, \ldots, r; (\cdot - x)_+^{r-1}],$$

where $[s_0, \ldots, s_r; f]$ denotes the divided difference of f taken at the points $s_0, \ldots, s_r$. Namely,

$$N_{i,h,\theta}^{(r)}(x) = N^{(r)}\left(\frac{x-t_i}{h}\right).$$

It is well known that

$$\sum_{i \in Z} N_{i,h,\theta}^{(r)}(x) = 1 \quad \text{for } x \in R, \tag{2.2}$$

where $R = (-\infty, \infty)$. For a later convenience we introduce

$$M_{i,h,\theta}^{(r)}(x) = h^{-1} N_{i,h,\theta}^{(r)}(x).$$

Now, since

$$\int_R N^{(r)}(x)dx = 1,$$

it follows that

$$\int_R M_{i,h,\theta}^{(r)}(x)dx = 1. \tag{2.3}$$

For the derivative of the cardinal B-spline we have the formula

$$DN^{(r+1)}(x) = M^{(r)}(x) - M^{(r)}(x-1) = N^{(r)}(x) - N^{(r)}(x-1), \tag{2.4}$$

whence

$$DN_{i,h,\theta}^{(r+1)} = M_{i,h,\theta}^{(r)} - M_{i+1,h,\theta}^{(r)}. \tag{2.5}$$

Proposition 2.6. *Let $1 \leqslant m < r$ and let $g = \sum_i a_i N_{i,h,\theta}^{(r)}$. Then*

$$h^m D^m g = \sum_i (\Delta^m a_i) N_{i+m,h,\theta}^{(r-m)} \tag{2.7}$$

and

$$\Delta_g^m g = \sum_i (\Delta^m a_i) N_{i,h,\theta}^{(r)}, \tag{2.8}$$

where $\Delta a_i = a_{i+1} - a_i$, $\Delta_h g(x) = g(x+h) - g(x)$ *and* $\Delta^{m+1} = \Delta^m(\Delta)$.

Proof. It is sufficient to apply (2.5) and the definition of $N_{i,h,\theta}^{(r)}$.

In what follows, for given positive functions $f(t)$, $t \in T$, and $g(t)$, $t \in T$, we write $f \sim g$ whenever there are two positive constants C_1, C_2 such that the inequalities $C_1 g(t) \leqslant f(t) \leqslant C_2 g(t)$ hold for all $t \in T$.

The following result is given in [12]:

PROPOSITION 2.9. *Let* $g = \sum_i a_i N_{i,h,\theta}^{(r)}$ *and let* $1 \leqslant p \leqslant \infty$. *Then* $\|g\|_p \sim h^{1/p} \|a\|_p$, *where the constants in these inequalities depend only on* r, $\|a\|_p = (\sum_i |a_i|^p)^{1/p}$ *for finite* p *and* $\|a\|_\infty = \sup\{|a_i|: i \in Z\}$.

In order to state the next result we recall the definition of the modulus of smoothness of order m of $f \in L^p(R)$ in the L^p-norm, $1 \leqslant p \leqslant \infty$,

$$\omega_{m,p}(f; \delta) = \sup_{|t| < \delta} \|\Delta_t^m f\|_p, \tag{2.10}$$

where Δ_t^m is the m-th order progressive difference with step t, i.e.,

$$\Delta_t^m f(x) = \sum_{j=0}^{m} (-1)^{j+m} \binom{m}{j} f(x+jt),$$

and

$$\|f\|_p = \begin{cases} \left(\int_R |f(x)|^p dx\right)^{1/p} & \text{if } 1 \leqslant p < \infty, \\ \operatorname{ess\,sup}\{|f|: x \in R\} & \text{if } p = \infty. \end{cases}$$

COROLLARY 2.11. *Let* $1 \leqslant m < r$, $1 \leqslant p \leqslant \infty$, *and let* $g = \sum_i a_i N_{i,h,\theta}^{(r)}$. *Then, for* $h > 0$,

$$\omega_{m,p}(g; h) \sim h^m \|D^m g\|_p \sim \|\Delta_h^m g\|_p \sim h^{1/p} \left(\sum_i |\Delta^m a_i|^p\right)^{1/p},$$

and the constants in these inequalities depend on r *only.*

Proof. Apply Propositions 2.9 and 2.6, definition (2.10) and the inequality

$$\omega_{m,p}(g; h) \leqslant h^m \|D^m g\|_p.$$

It is well known and it follows by induction with the help of (2.2) and (2.5) that the nontrivial B-splines over a given interval are linearly independent over that interval. In particular, every polynomial of degree not exceeding $r-1$ is a linear combination of the B-splines. In addition, we have

$$(y-x)^{r-1} = \sum_i \phi_{i,r,\theta}(y) N_{i,h,\theta}^{(r)}(x) \quad \text{for } x, y \in R,$$

where

$$\phi_{i,r,\theta}(y) = (y-t_{i+1})\dots(y-t_{i+r-1}), \qquad \phi_{i,1,\theta}(y) = 1.$$

Comparing the coefficients on both sides of the last but one formula at the powers of y gives for $0 \leqslant m < r$

$$x^m = \sum_i \left(\frac{1}{\binom{r-1}{m}} \sum_{0<j_1<\dots<j_m<r} t_{i+j_1}\dots t_{i+j_m} \right) N_{i,h,\theta}^{(r)}(x). \tag{2.12}$$

It is also important to recall the formula for the moments of the spline density $M_{j,h,\theta}^{(r)}$. Namely, for a nonnegative integer m we have

$$\int_R M_{i,h,\theta}^{(r)}(y) y^m dy = \frac{1}{\binom{r+m}{m}} \sum_{0\leqslant j_1\leqslant\dots\leqslant j_m\leqslant r} t_{i+j_1}\dots t_{i+j_m}. \tag{2.13}$$

In our estimation method the following properties of the spline densities and distributions are important. For a given spline probability density g such that

$$g(x) = \sum_i g_i N_{i,h,\theta}^{(r)}(x), \tag{2.14}$$

we obtain by (2.5) for the probability distribution function

$$G(x) = \int_{-\infty}^{x} g(y)dy$$

the formula

$$G(x) = \sum_i G_i N_{i,h,\theta}^{(r+1)}(x), \tag{2.15}$$

where

$$G_i = h \sum_{j\leqslant i} g_j. \tag{2.16}$$

It is important that the values $g(x)$ and $G(x)$ can easily be calculated by the numerically stable algorithm due to C. de Boor, M. G. Cox and L. Mansfield (see, e.g., [2]).

3. The direct approximation theorem by local spline operators. The approximating operators can now be defined for the integers $k > 0$, $r > 0$ and for any $f \in L_{\mathrm{loc}}^1(R)$ as follows:

$$Q_{h,\theta}^{(r,k)}(f;x) = \sum_{i\in Z} (f, M_{i+1,h,\theta}^{(k)}) N_{i,h,\theta}^{(r)}(x), \tag{3.1}$$

where l is a nonnegative integer such that $k+2l = r$ and

$$(f, g) = \int_R f(x)g(x)dx.$$

A pair (r, k) of positive integers is called *admissible* if $r \geqslant k$ and $r-k$ is even. Since the coefficient functional (the restriction $r \geqslant k$ is not essential) $f \mapsto (f, M^{(k)}_{i+l,h,\theta})$ is local for admissible (k, l) and $h > 0$, i.e., its support is contained in the support of $N^{(r)}_{i,h,\theta}$, these operators are *local spline operators* in de Boor's terminology [2]. These operators are not projections except for the case $r = k = 1$ (see [8]) but their L^p-norms are equal to 1, i.e., for all $1 \leqslant p \leqslant \infty$

$$\|Q^{(r,k)}_{h,\theta}(f)\|_p \leqslant \|f\|_p \quad \text{for } f \in L^p(R). \tag{3.2}$$

It also follows by (3.1), (2.2) and (2.3) that these operators have in addition the following properties:

1° $Q^{(r,k)}_{h,\theta}(f) \geqslant 0$ for $f \geqslant 0$.

2° $Q^{(r,k)}_{h,\theta}(1) = 1$.

3° For $f \in L^1(R)$

$$\int_R f = 1 \text{ implies } \int_R Q^{(r,k)}_{h,\theta}(f) = 1.$$

Consequently, the operator $Q^{(r,k)}_{h,\theta}$: $L^1(R) \to L^1(R)$ takes probability densities into probability densities, and its kernel will be used for constructing nonparametric density estimators.

PROPOSITION 3.3. *Let (r, k) be admissible. Then for any linear function f_0 we have*

$$Q^{(r,k)}_{h,\theta}(f_0) = f_0.$$

Proof. Since the operator has property 2°, it is enough to check the statement for $f_0(x) = x$, but this is implied by formulas (2.12) and (2.13) with $m = 1$.

To prove the main direct approximation theorem the following is needed:

LEMMA 3.4. *Let (r, k) be admissible. Moreover, let the function f be defined on R and let it satisfy the Lipschitz condition*

$$\omega_{1,\infty}(f; \delta) \leqslant \delta \quad \textit{for } \delta > 0.$$

Then

$$\|f - Q^{(r,k)}_{h,\theta}(f)\|_\infty \leqslant \frac{r+k}{2} h \quad \textit{for } h > 0.$$

Proof. For given $x \in R$ let i be such that $t_i < x < t_{i+1}$. Now,

$$f(x) - Q_{h,\theta}^{(r,k)}(f; x) = \sum_{i-r<j\leqslant i} N_{j,h,\theta}^{(r)}(x) \int_{t_{j+l}}^{t_{j+l+k}} M_{j+l,h,\theta}^{(k)}(y)\big(f(x)-f(y)\big)dy,$$

$$|f(x) - Q_{h,\theta}^{(r,k)}(f; x)| \leqslant \sum_{i-r<j\leqslant i} N_{j,h,\theta}^{(r)}(x) \int_{t_{j+l}}^{t_{j+l+k}} M_{j+l,h,\theta}^{(k)}(y)|x-y|dy$$

$$\leqslant \frac{r+k}{2}h \sum_{i-r<j\leqslant i} N_{j,h,\theta}^{(r)}(x) \int_{t_{j+l}}^{t_{j+l+k}} M_{j+l,h,\theta}^{(k)}(y)dy = \frac{r+k}{2}h.$$

In what follows $W_p^m(R)$ and $W_{p,\mathrm{loc}}^m(R)$ denote the usual Sobolev spaces over R with the index of smoothness m and the exponent of integration p.

The next lemma was earlier applied in [1] in the case of Bernstein polynomials and as appears it is natural to use the same idea in the case of our local spline operators.

LEMMA 3.5. *Let* $f \in W_p^2(R)$ *and let* $\operatorname{supp} D^2 f \subset [a, b]$ *for some finite* $a, b \in R$. *Then for all* $x \in R$ *we have*

$$f(x) = \frac{b-x}{b-a} f(a) + \frac{x-a}{b-a} f(b) - \int_a^b K_{a,b}(x, y) D^2 f(y) dy,$$

where $D = d/dx$ *and*

$$K_{a,b}(x, y) = \frac{1}{b-a} \min(x-a, y-a) \min(b-x, b-y).$$

Proof. Apply Peano's formula for the second order divided differences.

In the space $L^\infty(R)$ we consider its separable subspace $C_{-,+}(R)$ of all functions continuous with finite limits at $+\infty$ and at $-\infty$.

We are ready now to prove the basic direct result on the order of approximation by the local spline $Q_{h,\theta}^{(r,k)}$ operators.

THEOREM 3.6. *Let* (r, k) *be admissible with* $r > 1$. *Then, for* $f \in L^p(R)$ *in the case* $1 \leqslant p < \infty$ *and for* $f \in C_{-,+}(R)$ *in the case* $p = \infty$, *we have*

$$\|f - Q_{h,\theta}^{(r,k)}(f)\|_p \leqslant C_{r,k}\, \omega_{2,p}(f; h) \quad \text{for } h > 0,$$

where $C_{r,k} = 8 + 2(r+k)^2$. *Moreover, if* $f \in L_{\mathrm{loc}}^1(R)$ *and* $x \in R$ *is a strong Lebesgue point of* f, *then*

$$Q_{h,\theta}^{(r,k)}(f)(x) \to f(x) \quad \text{as } h \to 0_+.$$

Proof. Assume at first that $p < \infty$. Since the L^p-norms of the operators $Q_{h,\theta}^{(r,k)}$ are equal to 1 and $\omega_{2,p}(f; h) \leqslant 4\|f\|_p$, it is sufficient to prove the result for $f \in L^p(R)$ with compact support. Using Steklov means we define in a standard way the smoothing of f of order m, i.e.,

$$g(x) = -\sum_{j=1}^{m} (-1)^{j+m}\binom{m}{j}\int_0^1 \dots \int_0^1 f(x+jh(s_1+\dots+s_m))ds_1 \dots ds_m.$$

The function g is in $W_p^m(R)$. If the support of f is contained in $[a', b']$, then the support of g is contained in $[a, b]$ with $b' = b$ and $a = a' - m^2 h$. Moreover,

$$f(x) - g(x) = \int_0^1 \dots \int_0^1 \Delta^m_{h(s_1+\dots+s_m)} f(x) ds_1 \dots ds_m,$$

whence

$$\|f-g\|_p \leqslant m^m \omega_{m,p}(f; h). \tag{3.7}$$

It also follows that

$$D^m g(x) = -\sum_{j=1}^{m} (-1)^{j+m} \frac{1}{(jh)^m}\binom{m}{j} \Delta^m_{jh} f(x).$$

Thus

$$h^m \|D^m g\|_p \leqslant 2^m \omega_{m,p}(f; h). \tag{3.8}$$

In what follows it is enough to take $m = 2$. The next step is to prove

$$\|g - Q_{h,\theta}^{(r,k)}(g)\|_p \leqslant \frac{(r+k)^2}{2} h^2 \|D^2 g\|_p. \tag{3.9}$$

The locality of the operators $Q_{h,\theta}^{(r,k)}$, Proposition 3.3 and Lemma 3.5 give

$$\begin{aligned} Q_{h,\theta}^{(r,k)}(g; x) - g(x) \\ = \int_{|x-y|<h(r+k)/2} \big(K_{a,b}(x, y) - Q_{h,\theta}^{(r,k)}(K_{a,b}(\cdot, y); x)\big) D^2 g(y) dy. \end{aligned} \tag{3.10}$$

Now, the function $K_{a,b}(\cdot, y)$ for fixed $y \in [a, b]$ satisfies a Lipschitz condition in the L^∞-norm with constant 1. Lemma 3.4 and (3.10) then give

$$|Q_{h,\theta}^{(r,k)}(g; x) - g(x)| \leqslant \frac{(r+k)^2}{2} h^2 \frac{1}{(r+k)h} \int_{|x-y|<h(r+k)/2} |D^2 g(y)| dy,$$

and this implies (3.9). Applying now the triangle inequality to the identity

$$f - Q_{h,\theta}^{(r,k)}(f) = (f-g) + (g - Q_{h,\theta}^{(r,k)}(g)) + Q_{h,\theta}^{(r,k)}(g-f)$$

we obtain

$$\|f - Q_{h,\theta}^{(r,k)}(f)\|_p \leqslant 2\|f-g\|_p + \|g - Q_{h,\theta}^{(r,k)}(g)\|_p. \tag{3.11}$$

Now, (3.11), (3.7), (3.9) and (3.8) give the inequality stated in the theorem. In the case $p=\infty$ we reduce the proof to $f\in W^1_{\infty,\mathrm{loc}}(R)$ with finite supp Df, and then proceed like in the previous case. The proof of the convergence at the Lebesgue point is standard and it is omitted.

Since we are interested in estimating simultaneously the densities and the probability distribution functions, it is convenient to introduce the operators

$$T^{(r,k)}_{h,\theta}(F;x)=\sum_{i\in Z}\int_R M^{(k)}_{i+l,h,\theta}dF\int_{-\infty}^{x}N^{(r)}_{i,h,\theta}(y)dy. \tag{3.12}$$

PROPOSITION 3.13. *Let (r,k) be admissible. Then*

$$T^{(r,k)}_{h,\theta}\colon\ C_{-,+}(R)\to C_{-,+}(R), \tag{3.14}$$

and it preserves the limits at $+\infty$ and at $-\infty$. For F bounded and defined everywhere on R we have

$$\|T^{(r,k)}_{h,\theta}(F)\|_\infty\leqslant\|F\|_\infty, \tag{3.15}$$

where the $\|\cdot\|_\infty$-norm in the case $k=1$ has to be replaced by the customary max *norm. Now, if $k>1$, then for $F\in C_{-,+}(R)$ and $h>0$ we have*

$$\|F-T^{(r,k)}_{h,\theta}(F)\|_\infty\leqslant 2\big(4+(r+k)^2\big)\omega_{2,\infty}(F;h), \tag{3.16}$$

and if $k=1$, then

$$\|F-T^{(r,k)}_{h,\theta}(F)\|_\infty\leqslant\omega_{1,\infty}\left(F;\frac{r+k}{2}h\right). \tag{3.17}$$

Proof. Let $r-k=2l$. Note that for $k>1$

$$T^{(r,k)}_{h,\theta}(F)=Q^{(r+1,k-1)}_{h,\theta}(F), \tag{3.18}$$

and therefore (3.16) follows from Theorem 3.6. The proof of (3.17) is direct. Namely,

$$T^{(r,k)}_{h,\theta}(F)=\sum_{i\in Z}F(t_{i+l+1})N^{(r+1)}_{i,h,\theta}\qquad\text{with } k=1,\ l=(r-1)/2. \tag{3.19}$$

Now, for $x\in(t_j,t_{j+1})$ we get

$$T^{(r,k)}_{h,\theta}(F;x)-F(x)=\sum_{j-r<i\leqslant j}\big(F(t_{i+l+1})-F(x)\big)N^{(r+1)}_{i,h,\theta}(x), \tag{3.20}$$

whence (3.17) follows.

PROPOSITION 3.21. *Let (r,k) be admissible and let F be a function on R of bounded total variation on each finite interval. Then at each continuity point x of F*

$$T^{(r,k)}_{h,\theta}(F;x)\to F(x)\qquad \text{as } h\to 0_+.$$

Proof. In the case $k>1$ we use (3.18) and apply Theorem 3.6. For $k=1$ the statement follows directly from (3.20).

4. The saturation theorem. The aim of this section is to show that the order of approximation in Theorem 3.6 cannot be improved by increasing the smoothness of the approximating function.

PROPOSITION 4.1. *Let (r, k) be admissible with $r > 1$. Then for a continuous concave function f on R we have the inequality*

$$f(x) \geqslant Q_{h,\theta}^{(r,k)}(f; x) \quad \textit{for } x \in R,\ h > 0.$$

Proof. Using the definition of $Q_{h,\theta}^{(r,k)}(f)$, (2.13) and (2.12) we find by Jensen's inequality that

$$Q_{h,\theta}^{(r,k)}(f; x) \leqslant \sum_i f\big(h(i+\theta+r/2)\big)N_{i,h,\theta}^{(r)}(x) \leqslant f\big(\sum_i h(i+\theta+r/2)N_{i,h,\theta}^{(r)}(x)\big) = f(x).$$

PROPOSITION 4.2. *Let (r, k) be admissible with $r > 1$ and let $f_0(x) = x^2$. Then*

$$\frac{Q_{h,\theta}^{(r,k)}(f_0; x) - f_0(x)}{h^2} = \frac{r+k}{12} \quad \textit{for } x \in R.$$

Proof. Proposition 3.3 implies that for the proof the modified $f_0(x) = (x-h\theta)^2$ can be used. According to (2.13), for $r = k+2l$ we have

$$\int_R M_{i+l,h,\theta}^{(k)}(y) f_0(y)dy = h^2\big((i+r/2)^2 + k/12\big).$$

Moreover, (2.12) gives

$$f_0 = \sum_i h^2\big((i+r/2)^2 - r/12\big)N_{i,h,\theta}^{(r)},$$

and this completes the proof.

We are now ready to prove the main result of this section.

THEOREM 4.3. *Let (r, k) be admissible with $r > 1$. Let $f \in W_{1,\mathrm{loc}}^2(R)$ and let $x \in R$ be the strong Lebesgue point of $D^2 f$. Then*

$$\lim_{h\to 0+} \frac{Q_{h,\theta}^{(r,k)}(f; x) - f(x)}{h^2} = \frac{r+k}{24} D^2 f(x).$$

Proof. For the proof let a, b be some finite numbers such that $A < x < B$, where $A = a+(b-a)/4$ and $B = b-(b-a)/4$. Let us define the function ψ of class $C^2(R)$ as follows:

$$\psi(x) = \begin{cases} 1 & \text{if } x \in (A, B), \\ 0 & \text{if } x \notin (a, b), \end{cases} \quad \psi(x) \geqslant 0 \text{ elsewhere in } R.$$

Moreover, let $g = f\psi$. Then by (3.10) we get

$$\frac{Q_{h,\theta}^{(r,k)}(g; x) - g(x)}{h^2} = \int_R L_h(x, y) D^2 g(y)dy, \tag{4.4}$$

where

$$L_h(x, y) = \frac{1}{h^2}\left[K_{a,b}(x, y) - Q_{h,\theta}^{(r,k)}(K_{a,b}(\cdot, y); x)\right].$$

The kernel L_h is an approximate identity. It has the following properties:

1° $L_h(x, y) \geqslant 0$ for $x, y \in R$.

2° $L_h(x, y) = 0$ for $|x-y| > h(r+k)/2$.

3° $L_h(x, y) \leqslant (r+k)/2h$ for $x, y \in R$.

4° For small $h > 0$ we have

$$\int_R L_h(x, y)dy = \frac{r+k}{24} \quad \text{for } x \in R.$$

Property 1° follows from Proposition 4.1. The locality of the operator $Q_{h,\theta}^{(r,k)}$ and Proposition 3.3 imply 2°, and Proposition 3.4 gives 3°. To see the last property we note that (4.4) implies

$$\frac{Q_{h,\theta}^{(r,k)}(f; x) - f(x)}{h^2} = \int_R L_h(x, y) D^2 f(y) dy \quad \text{for small } h > 0, \tag{4.5}$$

which in particular for $f(x) = f_0(x) = x^2$ in combination with Proposition 4.2 gives 4°.

Now, since the point x is strong Lebesgue, we have

$$\frac{1}{(r+k)h} \int_{|y-x| < h(r+k)/2} |D^2 f(y) - D^2 f(x)| dy = o(1) \quad \text{as } h \to 0_+.$$

Thus, properties 1°–4° and (4.5) give for $h \to 0_+$

$$\left|\frac{1}{h^2}[Q_{h,\theta}^{(r,k)}(f; x) - f(x)] - \frac{r+k}{24} D^2 f(x)\right| \leqslant \int_R L_h(x, y) |D^2 f(y) - D^2 f(x)| dy = o(1)$$

and this completes the proof.

COROLLARY 4.6. *Let (r, k) be admissible with $r > 1$. Assume that $f \in W_p^2(R)$ for some p, $1 \leqslant p < \infty$, or that $f \in C_{-,+}(R) \cap C^2(R)$ in the case $p = \infty$. Then*

$$\|f - Q_{h,\theta}^{(r,k)}(f)\|_p = o(h^2) \quad \text{as } h \to 0_+$$

implies that $f = 0$.

5. The inverse approximation theorem. For the given set of simple knots (2.1) we introduce the linear spaces

$$S_{h,\theta}^r(R) = \operatorname{span}\{N_{i,h,\theta}^{(r)}: i \in Z\},$$

for finite $p \geqslant 1$

$$S_{h,\theta}^{r,p}(R) = S_{h,\theta}^r(R) \cap L^p(R),$$

and

$$S_{h,\theta}^{r,\infty}(R) = S_{h,\theta}^r(R) \cap C_{-,+}(R).$$

For each function $f \in L^p(R)$ the best approximation by $S^r_{h,\theta}(R)$ is the number

$$\mathscr{E}^{(r;p)}_{h,\theta}(f) = \inf\{\|f-g\|_p\colon\; g \in S^{r;p}_{h,\theta}(R)\}.$$

The inverse approximation theorems are statements in which the moduli of smoothness $\omega_{m,p}(f;\, h)$ are estimated from above by some means of $\mathscr{E}^{(r;p)}_{h,\theta}(f)$. One way of handling the best approximations $\mathscr{E}^{(r;p)}_{h,\theta}(f)$ is comparing them with approximations by orthogonal projections. To this end we introduce the *orthogonal projection* $P^{(r)}_{h,\theta}$ of $L^2(R)$ onto $S^{r;2}_{h,\theta}(R)$ and write

$$P^{(r)}_{h,\theta}(f;\, x) = \int_R K^{(r)}_{h,\theta}(x,\, y) f(y)dy, \tag{5.1}$$

where $K^{(r)}_{h,\theta}$ is the uniquely determined kernel. We have proved in [4]

PROPOSITION 5.2. *The kernel $K^{(r)}_{h,\theta}$ for $h > 0$ and for $\theta \in R$ has the following properties*:

$$K^{(r)}_{h,\theta}(x,\, y) = K^{(r)}_{h,\theta}(y,\, x) \quad \textit{for } x,\, y \in R, \tag{5.3}$$

$$K^{(r)}_{h,\theta}(x,\, \cdot) \in S^{r;p}_{h,\theta}(R) \quad \textit{for } x \in R,\ 1 \leqslant p \leqslant \infty, \tag{5.4}$$

$$|K^{(r)}_{h,\theta}(x,\, y)| \leqslant C\cdot\frac{1}{h} q^{|x-y|/h} \quad \textit{for } x,\, y \in R, \tag{5.5}$$

where $C > 0$ and q, $0 < q < 1$, depend on r only. Moreover, we have

$$\int_R K^{(r)}_{h,\theta}(x,\, y) y^i dy = x^i \quad \textit{for } x \in R,\ i = 0,\, \ldots,\, r-1. \tag{5.6}$$

COROLLARY 5.7. *The operators $P^{(r)}_{h,\theta}\colon L^p(R) \to L^p(R)$ are projections onto $S^{r;p}_{h,\theta}(R)$. For each integer $r \geqslant 1$ there is a constant C_r depending on r only such that*

$$\|P^{(r)}_{h,\theta}(f)\|_p \leqslant C_r \|f\|_p \quad \textit{for } 1 \leqslant p \leqslant \infty,\ h > 0,\ \theta \in R. \tag{5.8}$$

To prove the inverse result the following well-known elementary inequality for progressive differences (see, e.g., [3]) is needed:

LEMMA 5.9. *Let the integers $i \geqslant 0, j \geqslant 0$ be given and let $f \in W^i_p(R)$ for some p, $1 \leqslant p \leqslant \infty$. Then*

$$\|\Delta^{i+j}_h f\|_p \leqslant h^i \|D^i \Delta^j_h f\|_p \quad \textit{for } h > 0. \tag{5.10}$$

Moreover, the following Bernstein type inequality for splines with equally spaced knots (cf. [3]) will be used:

LEMMA 5.11. *Let $r \geqslant 1$, $1 \leqslant p \leqslant \infty$, $\theta \in R$, $\delta > 0$ and $h > 0$ be given. Then there is $C_r < \infty$ depending on r only such that for $f \in S^{r;p}_{h,\theta}(R)$ we have for integer $m \geqslant 0$*

$$\|\Delta^m_\delta f\|_p \leqslant C_r (\delta/h)^m \|f\|_p \quad \textit{for } 0 \leqslant m < r \tag{5.12}$$

and

$$\|\Delta_\delta^r f\|_p \leqslant C_r(\delta/h)^{r-1+1/p}\|f\|_p. \tag{5.13}$$

COROLLARY 5.14. *The inequality in* (5.12) *in the case* $p = 1$ *holds true for* $m = r$ *as well.*

To state the main inverse result we need some more notation. In what follows we are interested in the approximation by splines corresponding to the parameter $\theta = r/2$ in (2.1). Let us introduce the number

$$\varkappa := \begin{cases} \frac{1}{2} & \text{if } r \text{ is even,} \\ \frac{1}{3} & \text{if } r \text{ is odd.} \end{cases} \tag{5.15}$$

We are now in a position to state the main result of this section. The proof follows the same argument as presented in [3] but it differs in some detail and for the sake of completeness it is outlined below.

THEOREM 5.16. *Let* $r \geqslant 1$, $1 \leqslant p \leqslant \infty$ *and let* θ *and* $\varkappa$ *be defined as above. Moreover, let* m *be an integer such that* $1 \leqslant m < r$ *and let* $\delta > 0$. *Then there is a finite constant* C_r *depending on* r *only such that*

$$\omega_{m,p}(f;\delta) \leqslant C_r\delta^m\Big(\|f\|_p + \sum_{i=0}^{N} \varkappa^{-im}\mathscr{E}^{(r,p)}_{\varkappa^i,\theta}(f)\Big), \tag{5.17}$$

$$\omega_{r,p}(f;\delta) \leqslant C_r\delta^{r-1+1/p}\Big(\|f\|_p + \sum_{i=0}^{N} \varkappa^{-i(r-1+1/p)}\mathscr{E}^{(r,p)}_{\varkappa^i,\theta}(f)\Big), \tag{5.18}$$

where $N = [\log_{1/\varkappa}(1/\delta)]$.

Proof. The argument will be outlined only in the case of finite p. Let us introduce

$$f_i = P^{(r)}_{\varkappa^i,\theta}(f).$$

Then

$$f = f_0 + \sum_{i=0}^{N-1}(f_{i+1} - f_i) + (f - f_N),$$

and by Lemma 5.11 with $h = 1$ we have

$$\|\Delta_\delta^m f_0\|_p \leqslant C_r'\delta^m\|f_0\|_p \leqslant C_r\delta^m\|f\|_p, \tag{5.19}$$

$$\|\Delta_\delta^r f_0\|_p \leqslant C_r'\delta^{r-1+1/p}\|f_0\|_p \leqslant C_r\delta^{r-1+1/p}\|f\|_p.$$

Now, the operators $P^{(r)}_{\varkappa^i,\theta}$ are, by (5.8), bounded projections onto $S^{r,p}_{\varkappa^i,\theta}$ and, by the choice of $\varkappa$ and θ we have made, it follows that $S_i \subset S_{i+1}$, where $S_i := S^{r,p}_{\varkappa^i,\theta}$.

Clearly, $f_i \in S_i$, and therefore by Lemma 5.11 we have

(5.20) $$\|\Delta_\delta^m(f_{i+1}-f_i)\|_p \leqslant C'_r(\delta/\varkappa^i)^m \|f_{i+1}-f_i\|_p \leqslant C_r(\delta/\varkappa^i)^m \mathscr{E}^{(r,p)}_{\varkappa^i,\theta}(f),$$
$$\|\Delta_\delta^r(f_{i+1}-f_i)\|_p \leqslant C'_r(\delta/\varkappa^i)^{r-1+1/p} \|f_{i+1}-f_i\|_p \leqslant C_r(\delta/\varkappa^i)^{r-1+1/p} \mathscr{E}^{(r,p)}_{\varkappa^i,\theta}(f).$$

Finally,

(5.21) $$\|f-f_N\|_p \leqslant C_r \mathscr{E}^{(r,p)}_{\varkappa^N,\theta}(f).$$

The combination of (5.19), (5.20) and of (5.21) gives (5.17) and (5.18).

We may now conclude this section with the basic characterization of Lipschitz classes in the L^1-norm by means of the orders of approximation by the operators $Q^{(r,k)}_{h,\theta}$, where $\theta = r/2$.

THEOREM 5.22. *Let $r \geqslant 2$, $0 < \alpha < 2$ and let $f \in L^1(R)$. Then the following conditions are equivalent*:

(5.23) $$\|f-Q^{(r,k)}_{h,\theta}(f)\|_1 = O(h^\alpha) \quad \text{as } h \to 0_+,$$

(5.24) $$\omega_{2,1}(f;\, h) = O(h^\alpha) \quad \text{as } h \to 0_+,$$

where (r, k) is an admissible pair. Moreover, for $\alpha = 2$ the relation (5.24) *implies* (5.23).

Proof. Using the obvious inequality

(5.25) $$\mathscr{E}^{(r,p)}_{h,\theta}(f) \leqslant \|f-Q^{(r,k)}_{h,\theta}(f)\|_p,$$

we find by Theorem 5.16 that (5.23) implies (5.24) for $\alpha < 2$. The remaining part is a consequence of Theorem 3.6.

6. The estimators and their consistency. We now assume that a probability space $(\Omega, \mathscr{F}, \Pr)$ is given. The mean value with respect to Pr is denoted by E. We also assume that a simple sample of size n, i.e., a sequence $(X_1, \ldots, X_n)$ of i.i.d. real-valued random variables, is given. Their common distribution is denoted by F. For the empirical distribution we use the standard notation

$$F_n(x) = \frac{|\{i\colon X_i < x,\ 1 \leqslant i \leqslant n\}|}{n}.$$

In what follows it is assumed that the pair (r, k) is admissible and that $\theta = r/2$.

DEFINITION 6.1. The estimator for $F(x)$ is defined as follows:

$$F^{(r,k)}_{n,h} = T^{(r,k)}_{h,\theta}(F_n) \quad \text{with } h > 0,$$

where the right-hand side is defined by (3.18) for $k > 1$ and by (3.19) for $k = 1$. If there will be no misunderstanding the upper index (r, k) at $F_{n,h}$ will be suppressed.

It follows immediately that

(6.2) $$EF_{n,h} = T^{(r,k)}_{h,\theta}(F_n;\, x) \quad \text{for } h > 0.$$

THEOREM 6.3. *Let F be a continuous probability distribution function on R. Then*

$$\Pr\{\|F-F_{n,h}\|_\infty \to 0 \text{ as } h\to 0_+,\, n\to\infty\} = 1.$$

Proof. For $k=1$, Definition 6.1, (6.2), (3.19) and (3.17) imply

$$\|F-F_{n,h}\|_\infty \leqslant \|F-T_{h,\theta}^{(r,k)}(F)\|_\infty + \|T_{h,\theta}^{(r,k)}(F)-T_{h,\theta}^{(r,k)}(F_n)\|_\infty \tag{6.4}$$

$$\leqslant \omega_{1,\infty}\left(F;\frac{r+k}{2}h\right)+\|F-F_n\|_\infty.$$

Since F is continuous, $\omega_{1,\infty}(F;h)\to 0$ as $h\to 0_+$ and by Glivenko's theorem we have

$$\Pr\{\|F-F_n\|_\infty\to 0 \text{ as } n\to\infty\}=1.$$

Similarly, for $k>1$ we obtain now, by (3.18), (3.17) and by (3.2) with $p=\infty$,

$$\|F-F_{n,h}\|_\infty \leqslant C_{r,k}\omega_{2,\infty}(F;h)+\|F-F_n\|_\infty, \tag{6.5}$$

and this completes the proof.

For a later use let $\mathscr{D}$ denote the set of all densities on R.

DEFINITION 6.6. For an absolutely continuous probability distribution function F let $f=DF$. The density estimator is now defined by the formula

$$f_{n,h}^{(r,k)} = DF_{n,h}^{(r,k)}.$$

It then follows that $f_{n,h}\in\mathscr{D}$. The upper index (r,k) will usually be suppressed unless it could cause a confusion.

PROPOSITION 6.7. *For the admissible (r,k) we have*

$$f_{n,h}(x) = \int_R Q_{h,\theta}^{(r,k)}(y,x)\,dF_n(y) \tag{6.8}$$

with

$$Q_{h,\theta}^{(r,k)}(y,x) = \sum_{i\in Z} M_{i+l,h,\theta}^{(k)}(y)N_{i,h,\theta}^{(r)}(x) \tag{6.9}$$

with $r=2l+k$ and for $h>0$. In particular,

$$Ef_{n,h}(x) = Q_{h,\theta}^{(r,k)}(f;x). \tag{6.10}$$

The simple proof is left out.

For the absolutely continuous probability distributions the natural distance is the *total variation* over R. Since

$$\|F-F_{n,h}\|_\infty \leqslant \mathrm{V}(F-F_{n,h};R) = \|f-f_{n,h}\|_1, \tag{6.11}$$

where $\mathrm{V}(g;R)$ is the total variation of g over R, the L^1-metric on the space $\mathscr{D}$ seems to be natural. For more arguments for the L^1-metric we refer to [9].

PROPOSITION 6.12. *Let $h > 0$ and let $\tau > 0$. Then there is a constant C_r, depending on r only, such that for any absolutely continuous F, $f = DF$, we have*

$$(6.13)\qquad \|f-f_{n,h}\|_1 \leqslant C_r\left(\|f-Q_{h,\theta}^{(r,k)}f\|_1+\frac{\tau}{h}\|F-F_n\|_\infty+\int_{|x|\geqslant\tau} f(x)\,dx\right).$$

Proof. Introducing $Qf = Q_{h,\theta}^{(r,k)}(f)$ we find

$$(6.14)\qquad \|f-f_{n,h}\|_1 \leqslant \|f-Qf\|_1+\|Qf-f_{n,h}\|_1.$$

We write the last term as follows:

$$(6.15)\qquad \|Qf-f_{n,h}\|_1 = \int_{|x|<\tau} |Qf-f_{n,h}|+\int_{|x|\geqslant\tau} |Qf-f_{n,h}|.$$

Using the fact that both Qf and $f_{n,h}$ are densities, for the latter term we obtain

$$(6.16)\qquad \int_{|x|\geqslant\tau} |Qf-f_{n,h}| \leqslant \int_{|x|<\tau} |Qf-f_{n,h}|+2\int_{|x|\geqslant\tau} Qf.$$

Moreover,

$$\int_{|x|\geqslant\tau} Qf \leqslant \int_{|x|\geqslant\tau} |Qf-f|+\int_{|x|\geqslant\tau} f \leqslant \|Qf-f\|_1+\int_{|x|\geqslant\tau} f.$$

This, (6.16), (6.15) and (6.14) give

$$(6.17)\qquad \|f-f_{n,h}\|_1 \leqslant 3\|Qf-f\|_1+2\int_{|x|<\tau} |Qf-f_{n,h}|+2\int_{|x|\geqslant\tau} f.$$

It now follows by the definitions of Qf and $f_{n,h}$ that $Qf-f_{n,h} = Dg$, where

$$g = T_{h,\theta}^{(r,k)}(F-F_n)\in S_{h,\theta}^{(r)}.$$

However, Proposition 2.9 and Corollary 2.11 imply the Bernstein type inequality

$$(6.18)\qquad \|Dg\|_p \leqslant C_r h^{-1}\|g\|_p \quad \text{for } g\in S_{h,\theta}^{(r)},$$

and therefore

$$\int_{|x|<\tau} |Qf-f_{n,h}| \leqslant 2\tau\|Dg\|_\infty \leqslant C_r\frac{\tau}{h}\|g\|_\infty.$$

Since, by (3.15), $\|g\|_\infty \leqslant \|F-F_n\|_\infty$, the proof of (6.13) now follows by (6.17).

THEOREM 6.19. *Let $r > 1$ and let f be a given density on R. Then*

$$\Pr\{\|f-f_{n,h}\|_1 \to 0 \text{ as } n\to\infty\} = 1$$

provided that h depends on n in such a way that

$$\frac{\log n}{n} = o(h^2) \quad \text{as } n\to\infty,\ h\to 0_+.$$

Proof. The argument is based on the inequality (6.13). Theorem 3.6 implies that

$$\|f - Q_{h,\theta}^{(r,k)}(f)\|_1 \to 0 \quad \text{as } h \to 0_+.$$

Since $f \in L^1(R)$, we have

$$\int_{|x| \geqslant \tau} f(x)dx \to 0 \quad \text{as } \tau \to \infty.$$

To estimate the middle term of the right-hand side of (6.13) we use the inequality

$$\Pr\left\{\|F - F_n\|_\infty > \frac{\lambda}{\sqrt{n}}\right\} \leqslant C\exp(-2\lambda^2), \quad \lambda > 0, \tag{6.20}$$

which follows by the well-known N. V. Smirnov formula for the distribution of

$$\sup\{F(x) - F_n(x)\colon\ x \in R\}$$

(see [11]). Substituting $\lambda = \varepsilon h\sqrt{n}$ in (6.20), we obtain

$$\Pr\{h^{-1}\|F - F_n\|_\infty > \varepsilon\} \leqslant C\exp\big(-2\varepsilon^2(h^2 n)\big). \tag{6.21}$$

Now, the elementary convergence properties of Dirichlet series imply

$$\Pr\{h^{-1}\|F - F_n\|_\infty \to 0 \text{ as } n \to \infty\} = 1 \tag{6.22}$$

whenever $(\log n)/n = o(h^2)$. Choosing

$$\tau^{-1} = \sqrt{h^{-1}\|F - F_n\|_\infty}$$

we see that $\tau \to \infty$ as $n \to \infty$. Thus

$$\Pr\left\{\frac{\tau}{h}\|F - F_n\|_\infty \to 0 \text{ as } n \to \infty\right\} = 1,$$

and this completes the proof.

Recently G. Krzykowski has proved Theorem 6.19 under the weaker condition: $nh \to \infty$ as $n \to \infty$ and $h \to 0_+$.

7. Asymptotic orders of approximation by estimators for distributions. It is assumed in this section that the dependence of h on n is such that

$$h \sim 1/n^\beta. \tag{7.1}$$

We also assume that θ is given as in the paragraph following Corollary 5.14.

THEOREM 7.2. *Let (r, k) be admissible and let $\alpha\beta < \frac{1}{2}$, where $0 < \alpha \leqslant 2$ and $\beta > 0$. Moreover, let F be a continuous probability distribution on R. Then, for $k > 1$,*

$$\omega_{2,\infty}(F; h) = O(h^\alpha) \quad \textit{as } h \to 0_+ \tag{7.3}$$

implies

$$\Pr\{\|F-F_{n,h}^{(r,k)}\|_\infty = O(1/n^{\alpha\beta}) \text{ as } n\to\infty\} = 1. \tag{7.4}$$

Conversely, for $0<\alpha<2$, (7.4) *implies* (7.3). *In the case* $k=1$ *and* $0<\alpha\leqslant 1$ *the condition*

$$\omega_{1,\infty}(F;h) = O(h^\alpha) \quad \text{as } h\to 0_+ \tag{7.5}$$

implies (7.4). *Conversely, for* $0<\alpha<1$, (7.4) *implies* (7.5).

Proof. Let $k>1$. Assuming $0<\alpha\leqslant 2$, by (7.3) and (6.5) we obtain

$$\|F-F_{n,h}\|_\infty \leqslant C\frac{1}{n^{\alpha\beta}}+\|F-F_n\|_\infty. \tag{7.6}$$

Now, (6.21) and (6.22) with $h=(\log n)/\sqrt{n}$ give with probability 1

$$\|F-F_n\|_\infty = o\big((\log n)/\sqrt{n}\big). \tag{7.7}$$

Therefore the combination of (7.6) and (7.7) implies (7.4). Conversely, let $0<\alpha<2$; then by the definition of the best approximation we have

$$\mathscr{E}_{h,\theta}^{(r,\infty)}(F) \leqslant \|F-F_{n,h}\|_\infty. \tag{7.8}$$

Now, (7.4) and (7.8) imply

$$\mathscr{E}_{h,\theta}^{(r,\infty)}(F) = O(h^\alpha), \tag{7.9}$$

whence by (5.17) with $m=2$ and $p=\infty$ we obtain (7.3). Let now $k=1$, $0<\alpha\leqslant 1$, and assume that (7.5) is satisfied. Then (6.4) and (7.7) imply (7.4). Conversely, if $0<\alpha<1$ and (7.4) holds, then by (7.8) and by (5.17) with $p=\infty$, $m=1$, we obtain (7.5).

THEOREM 7.10. *Let* (r,k) *be admissible and let* $\alpha\beta=\frac{1}{2}$, *where* $0<\alpha\leqslant 2$ *and* $\beta>0$. *Moreover, let F be a continuous distribution on R. Then, for* $k>1$, (7.3) *implies*

$$E\|F-F_{n,h}^{(r,k)}\|_\infty = O\big(1/\sqrt{n}\big) \quad \text{as } n\to\infty. \tag{7.11}$$

Conversely, for $0<\alpha<2$, (7.11) *implies* (7.3).

In the case $k=1$ *and* $0<\alpha\leqslant 1$, (7.5) *implies* (7.11). *Conversely, for* $0<\alpha<1$, (7.11) *implies* (7.5).

Proof. Let $k>1$ and $0<\alpha\leqslant 2$. Inequalities (7.6) and (6.20) give (7.11). Assuming (7.11) we find that (7.8) implies (7.9), whence by the same argument as in the previous proof (7.3) follows. Suppose now that $0<\alpha<2$. Then (7.8) and (7.11) imply (7.9), whence by (5.17) with $m=2$ and $p=\infty$ we obtain (7.3). If $k=1$, $0<\alpha\leqslant 1$ and if (7.5) holds, then (6.4) and (6.20) imply (7.11). Conversely, if $0<\alpha<1$ and if (7.11) holds, then by (7.8) and by (5.17) with $p=\infty$, $m=1$, we obtain (7.5).

COROLLARY 7.12. *Let (r, k) be admissible. Then for any probability distribution F with bounded density we have* (7.11) *with $h^2 \sim 1/n$, i.e., $\beta = \frac{1}{2}$. Moreover, this is the best possible result for a fairly large subclass of distributions with bounded densities.*

COROLLARY 7.13. *Let (r, k) be admissible and let $k > 1$. Let the density $f = DF$ satisfy the Lipschitz condition with exponent* 1. *Then* (7.11) *holds with $h^4 \sim 1/n$, i.e., $\beta = \frac{1}{4}$.*

EXAMPLE. Let (r, k) be admissible. The arcsin law is defined as follows:

$$F_0 = \begin{cases} 0 & \text{for } x \leqslant 0, \\ (2/\pi)\arcsin\sqrt{x} & \text{for } 0 < x < 1, \\ 1 & \text{for } x \geqslant 1. \end{cases}$$

One checks easily that

$$\omega_{2,\infty}(F_0; h) \leqslant 2\omega_{1,\infty}(F_0; h) \leqslant C\sqrt{h},$$

and therefore (7.11) takes place with $h = 1/n$, i.e., $\beta = 1$.

8. Asymptotic orders of approximation by estimators for densities. The estimators for $f \in \mathscr{D}$ constructed in this section differ from those given by Definition 6.6 in that they have smaller support controlled by the *tail* of the corresponding probability distribution.

For a given probability distribution F we define the *tail function* as follows:

$$\Phi_F(\lambda) = 1 - F(\lambda) - F(-\lambda) \quad \text{for } \lambda > 0. \tag{8.1}$$

For F absolutely continuous we set $f = DF$ and

$$\phi_f(\lambda) := \Phi_F(\lambda) = \int_{|x| \geqslant \lambda} f(y)\,dy. \tag{8.2}$$

Along with the operators $Q_h := Q_{h,\theta}^{(r,k)}$ we now consider

$$Q_{h,\lambda} f := Q_{h,\lambda,\theta}^{(r,k)}(f) = \sum_{-\lambda' \leqslant t_i \leqslant \lambda' - rh} (f, M_{i+l,h,\theta}^{(k)}) N_{i,h,\theta}^{(r)}, \tag{8.3}$$

where $\lambda' = \lambda + h(r+k)/2$, (r, k) is admissible, and $r = k + 2l$.

PROPOSITION 8.4. *Let (r, k) be admissible, $\lambda' = \lambda + h(r+k)/2$ and let $f \in \mathscr{D}$. Then*

1° $\operatorname{supp}(Q_{h,\lambda} f) \subset \langle -\lambda', \lambda' \rangle$,
2° $1 - \phi_f(\lambda) \leqslant \|Q_{h,\lambda} f\|_1 \leqslant 1$,
3° $\|Q_h f - Q_{h,\lambda} f\|_1 \leqslant \phi_f(\lambda)$.

The easy proof is left out.

In general, for $f \in \mathscr{D}$ the function $Q_{h,\lambda} f$ is not necessarily a density. This leads to a new density estimator.

We recall that $\theta = r/2$.

DEFINITION 8.5. For admissible (r, k) and for a simple sample $X_1, \ldots, X_n$ corresponding to the distribution F we define

$$f_{n,h,\lambda} := f_{n,h,\lambda}^{(r,k)} = \sum_{-\lambda' \leqslant t_i \leqslant \lambda' - rh} \Big(\int_R M_{i+l,h,\theta}^{(k)}\, dF_n\Big) N_{i,h,\theta}^{(r)}, \tag{8.6}$$

where $r = k+2l$, $\lambda' = \lambda+h(r+k)/2$, $h > 0$ and $\lambda > 0$. Moreover, let

$$f^*_{n,h,\lambda} = \frac{f_{n,h,\lambda}}{\|f_{n,h,\lambda}\|_1}. \tag{8.7}$$

It follows by Proposition 8.4, 1°, that $\operatorname{supp}(f^*_{n,h,\lambda}) \subset \langle -\lambda', \lambda' \rangle$.

PROPOSITION 8.8. *Let (r, k) be admissible and let $r > 1$. Moreover, let h and λ be as in Definition 8.5. Then for some constant C, depending on r only, we have*

$$E\|f-f^*_{n,h,\lambda}\|_1 \leqslant C[\omega_{2,1}(f; h)+1/\sqrt{n}+\sqrt{\lambda/nh}+\phi_f(\lambda)]. \tag{8.9}$$

Proof. Using (8.6) and (8.7) we find that

$$\|f-f^*_{n,h,\lambda}\|_1 \leqslant \|f-f_{n,h,\lambda}\|_1+\Phi_{F_n}(\lambda). \tag{8.10}$$

Moreover, if we use (8.3), then

$$\|f-f_{n,h,\lambda}\|_1 \leqslant \|f-Q_h f\|_1+\|Q_h f-Q_{h,\lambda} f\|_1+\|f_{n,h,\lambda}-Ef_{n,h,\lambda}\|_1. \tag{8.11}$$

Now,

$$\|Q_h f-Q_{h,\lambda} f\|_1 \leqslant \Phi_F(\lambda) = \phi_f(\lambda), \tag{8.12}$$

and

$$|\Phi_F(\lambda)-\Phi_{F_n}(\lambda)| \leqslant 2\|F-F_n\|_\infty. \tag{8.13}$$

Thus, combining (8.13), (8.12), (8.11) and (8.10) we get

$$\|f-f^*_{n,h,\lambda}\|_1 \leqslant \|f-Q_h f\|_1+2\|F-F_n\|_\infty+2\phi_f(\lambda)+\|f_{n,h,\lambda}-Ef_{n,h,\lambda}\|_1. \tag{8.14}$$

We estimate the mean value of the last term as follows:

$$\begin{aligned}
E\|f_{n,h,\lambda}-Ef_{n,h,\lambda}\|_1 &= 2\lambda'\frac{1}{2\lambda'}\int_{-\lambda'}^{\lambda'} E|f_{n,h,\lambda}(x)-Ef_{n,h,\lambda}(x)|\,dx \\
&\leqslant 2\lambda'\left(\frac{1}{2\lambda'}\int_{-\lambda'}^{\lambda'} E|f_{n,h,\lambda}(x)-Ef_{n,h,\lambda}(x)|^2\,dx\right)^{1/2} \\
&= 2\lambda'\left(\frac{1}{2\lambda'}\int_{-\lambda'}^{\lambda'} E\left|\frac{1}{n}\sum_{j=1}^{n}\left(Q_{h,\lambda}(X_j, x)-EQ_{h,\lambda}(X_j, x)\right)\right|^2 dx\right)^{1/2} \\
&\leqslant \sqrt{\frac{2\lambda'}{n}}\left(\int_{-\lambda'}^{\lambda'}\int_R Q^2_{h,\lambda}(y, x)f(y)\,dy\,dx\right)^{1/2} \\
&\leqslant \sqrt{\frac{2\lambda'}{nh}}\left(\int_{-\lambda'}^{\lambda'}\int_R Q_{h,\lambda}(y, x)f(y)\,dy\,dx\right)^{1/2} \\
&\leqslant \sqrt{\frac{2\lambda'}{nh}} \leqslant \sqrt{\frac{2\lambda}{nh}}+\sqrt{\frac{r+k}{n}},
\end{aligned}$$

where

$$Q_{h,\lambda}(y, x) = \sum_{-\lambda' \leqslant t_i \leqslant \lambda' - rh} M^{(k)}_{i+l,h,\theta}(y) N^{(r)}_{i,h,\theta}(x).$$

This, (8.14), (6.21) and Theorem 3.6 give (8.9).

THEOREM 8.15. *Let (r, k) be admissible and let $r > 1$. Moreover, let the parameters α and δ be given such that $0 < \alpha \leqslant 2$, $\delta \geqslant 0$. Define*

$$\beta = \frac{\delta}{\delta+\alpha+2\delta\alpha} \quad \text{and} \quad \gamma = \frac{\alpha}{\delta+\alpha+2\delta\alpha}.$$

Then for $f \in \mathscr{D}$ the conditions

$$\omega_{2,1}(f; h) = O(h^{\alpha}) \quad \text{as } h \to 0, \tag{8.16}$$

$$\phi_f(\lambda) = O(1/\lambda^{\delta}) \quad \text{as } \lambda \to \infty \tag{8.17}$$

imply

$$E\|f - f^*_{n,h,\lambda}\|_1 = O(1/n^{\alpha\beta}) \quad \text{as } n \to \infty, \tag{8.18}$$

where $h \sim 1/n^{\beta}$, and $\lambda \sim n^{\gamma}$. Conversely, for $0 < \alpha < 2$, $\delta > 0$ the property (8.18) *implies both* (8.16) *and* (8.17).

Proof. Since $\alpha\beta = \delta\gamma < \frac{1}{2}$, the direct statement follows by Proposition 8.8. The converse is obtained as follows. The definition of $\mathscr{E}^{r,1}_{h,\theta}$ and (8.18) give

$$\mathscr{E}^{r,1}_{h,\theta} \leqslant E\|f - f^*_{n,h,\lambda}\|_1 \leqslant C\frac{1}{n^{\alpha\beta}} \leqslant C'h^{\alpha},$$

whence the inequality (8.16) for $0 < \alpha < 2$ follows by (5.17) with $k = 2$ if $r > 2$ and by (5.18) in the case $r = 2$. To obtain the inequality (8.17) from (8.18) observe that

$$E\|f - f^*_{n,h,\lambda}\|_1 \geqslant E\int_{|x| > \lambda'} |f - f^*_{n,h,\lambda}| = \int_{|x| > \lambda'} f,$$

whence

$$\int_{|x| > \lambda/\varkappa} f \leqslant \int_{|x| > \lambda'} f = O(1/n^{\alpha\beta}) = O(1/n^{\delta\gamma}) = O(\lambda^{-\delta}).$$

COROLLARY 8.19. *Let the support of $f \in \mathscr{D}$ be bounded and let the condition* (8.16) *be satisfied. Then $\beta = 1/(1+2\alpha)$ ($\delta = \infty$) and there is an $s_0 \in Z$ determined by the size of the* $\operatorname{supp} f$ *such that*

$$E\|f - f^*_{n,m,s_0}\|_1 = O(n^{-\alpha/(1+2\alpha)}). \tag{8.20}$$

The case of $\alpha = 2$ when $\alpha/(1+2\alpha) = \frac{2}{5}$ (see [9], p. 37) is particularly interesting. Moreover, if the density f has not a bounded support but it rapidly decreases at ∞ to 0 and at $-\infty$ and if it satisfies (8.16), then for each $\varepsilon > 0$

there is $\delta > 0$ such that

(8.21) $$E\|f - f^{*}_{n,h,\lambda}\|_1 = O(n^{-\alpha/(1+2\alpha)+\varepsilon}).$$

Note that the condition (8.16) is satisfied for large classes of densities:

1° If $f \in W^1_{1,\mathrm{loc}}(R)$ and Df is of bounded variation on R, then (8.16) holds with $\alpha = 2$.

2° If f is of bounded variation on R, then (8.16) holds with $\alpha = 1$.

EXAMPLES. (1) (8.21) applies to the Gaussian law with $\alpha = 2$.

(2) (8.21) with $\alpha = 1$ applies to the exponential (and double exponential) law.

(3) (8.20) with $\alpha = \frac{1}{2}$ applies to the arcsin law.

(4) For the stable law (including nonsymmetric) with exponent μ, $0 < \mu < 2$, we have $\alpha = 2$ and $\delta = 1+\mu$. In this case (8.18) can be applied with $\alpha\beta = (2+2\mu)/(7+5\mu)$.

9. The optimal width of the window parameter. For a given simple sample $(X_1, \ldots, X_n)$ we repeat Definitions 6.1 and 6.6. Namely, for admissible (r, k) we have

(9.1) $$F_{n,h} := F^{(r,k)}_{n,h} = T^{(r,k)}_{h,\theta}(F_n) \quad \text{for } h > 0,$$

(9.2) $$f_{n,h} := f^{(r,k)}_{n,h} = DF^{(r,k)}_{n,h} \quad \text{for } h > 0.$$

The aim of this section is to derive the density estimator called a *window function*, i.e., a function of variables $h, X_1, \ldots, X_n$ which for a fixed sample will give the possibility of finding the *optimal* h by looking at its infimum.

PROPOSITION 9.3. *Assume that (r, k) is admissible and that $k > 1$. Then there is a constant C depending on r only such that*

(9.4) $$\mathrm{V}(F - F_{n,h}; R) \leqslant C[\omega_{2,1}(f; h) + h^{-1}\|T_{h,\theta}F - F_{n,h}\|_1], \qquad h > 0,$$

where $\mathrm{V}(g; R)$ is the total variation of g over R.

Proof. The Bernstein type inequality (5.12) and Theorem 3.6 give

$$\mathrm{V}(F - F_{n,h}; R) = \|D(F - F_{n,h})\|_1 \leqslant \|D(F - T_{h,\theta}F)\|_1 + \|D(T_{h,\theta}F - F_{n,h})\|_1$$

$$= \|f - Q_{h,\theta}f\|_1 + \|D(T_{h,\theta}F - F_{n,h})\|_1 \leqslant C[\omega_{2,1}(f; h) + h^{-1}\|T_{h,\theta}F - F_{n,h}\|_1].$$

In the right-hand side of (9.4) we look at the second term. In that term, F is unknown and we replace it by the known estimator $F_{n,h}$. Similarly, in the first term of the right-hand side of (9.4) the f is the unknown and we replace it by the known estimator $f_{n,h}$. Thus we end up with the following quantity:

(9.5) $$\omega_{2,1}(f_{n,h}; h) + h^{-1}\|T_{h,\theta}F_{n,h} - F_{n,h}\|_1.$$

We proceed with majorizing of this quantity. According to (3.18) and to Theorem 3.6 we have

$$h^{-1}\|T_{h,\theta}F_{n,h}-F_{n,h}\|_1 = h^{-1}\|Q_{h,\theta}^{r+1,k-1}F_{n,h}-F_{n,h}\|_1$$
$$\leqslant Ch^{-1}\omega_{2,1}(F_{n,h};\, h) \leqslant C\omega_{1,1}(f_{n,h};\, h),$$

and the first term in (9.5) is being estimated with the help of the trivial inequality $\omega_{2,1}(f_{n,h};\, h) \leqslant 2\omega_{1,1}(f_{n,h};\, h)$. Thus, up to a multiplicative constant, (9.5) is majorized by $\omega_{1,1}(f_{n,h};\, h)$. However, by Corollary 2.11 we have

$$\omega_{1,1}(f_{n,h};\, h) \sim h\sum_i |\Delta a_i|, \tag{9.6}$$

where the a_i's are given by the formula

$$f_{n,h} = \sum_i a_i N_{i,h,\theta}^{(r)}$$

or, more explicitly, according to (6.8)

$$a_i = n^{-1}\sum_{j=1}^{n} M_{i+l,h,\theta}^{(k)}(X_j).$$

Thus the right-hand side of (9.6), by using (2.5), can be written as

$$\sum_i \Big|n^{-1}\sum_{j=1}^{n} g^{(k)}(X_j/h-i)\Big|,$$

and therefore it is useful to introduce the function

$$\mathscr{F}^{(k)}(h;\, \boldsymbol{x}) := \sum_i \Big|n^{-1}\sum_{j=1}^{n} g^{(k)}(x_j/h-i)\Big|, \tag{9.7}$$

where $\boldsymbol{x} = (x_1, \ldots, x_n) \in R^n$ and

$$g^{(k)}(x) = N^{(k)}(x-k/2) - N^{(k)}(x+1-k/2). \tag{9.8}$$

Definition 9.9. For $k > 1$, $k \in Z$, $h > 0$ and for a given simple sample $X = (X_1, \ldots, X_n)$ we define the *window function* as follows:

$$\mathscr{F}(h) := \mathscr{F}^{(k)}(h;\, X), \tag{9.10}$$

where the right-hand side is given by formulas (9.7) and (9.8).

Since the window function is majorizing in the sense described above the error of estimating the distribution in the total variation norm or equivalently in estimating the density in the L^1-norm, we call the smallest $h_0 > 0$ at which $\mathscr{F}(h)$ attains its absolute minimum the *optimal width* of the window parameter h in the estimation process.

To state the main result on the window function we need the following periodic function:

$$G^{(k)}(x) = \sum_{i\in Z} |g^{(k)}(x-i)|. \tag{9.11}$$

Since $k > 1$, the function is continuous and it is also positive and periodic. We also introduce the constant

$$m_k = \min_{R} G^{(k)}. \tag{9.12}$$

It is also helpful to introduce the notation

$$\|x\|_\infty = \max(|x_1|, \ldots, |x_n|),$$
$$\delta(x) = \min\{|x_i - x_j|:\ x_i \neq x_j,\ i, j = 1, \ldots, n\},$$

where $x = (x_1, \ldots, x_n) \in R^n$.

PROPOSITION 9.13. *Let $k > 1$, $k \in Z$ and let $X \in R^n$ be a given simple sample. Then $\mathscr{F}(\cdot)$ is continuous, positive and bounded by 2 on the positive real axis. Moreover,*

$$\mathscr{F}(h) = \frac{1}{n}\sum_{j=1}^{n} G^{(k)}\left(\frac{X_j}{h}\right) \geqslant m_k \quad \textit{for } h \notin \left(\frac{\delta(X)}{k+1}, 2\|X\|_\infty\right), \tag{9.14}$$

and with probability one

$$\liminf_{h\to 0+} \mathscr{F}(h) = m_k. \tag{9.15}$$

Proof. For the function $g^{(k)}$ given by (9.6) we observe that

$$\operatorname{supp} g^{(k)} = \langle -k/2, k/2+1\rangle$$

and $\frac{1}{2}$ is the only zero of $g^{(k)}$ in $\langle -k/2, k/2+1\rangle$. Now, for fixed i, if h is such that $h > 2\|X\|_\infty$, then $X_j/h - i \in (-\frac{1}{2}-i, \frac{1}{2}-i)$ for $j = 1, \ldots, n$ and either $g^{(k)}(X_j/h - i) \geqslant 0$ for all j or $g^{(k)}(X_j/h - i) \leqslant 0$ for all j. Thus, $h > 2\|X\|_\infty$ implies (9.14). Let now

$$0 < h < \frac{\delta(X)}{k+1}.$$

Since $|\operatorname{supp} g^{(k)}| = k+1$, it follows that $X_j \neq X_{j'}$ implies that at least one of the numbers $g^{(k)}(X_j/h - i)$, $g^{(k)}(X_{j'}/h - i)$ is zero. In addition, if $X_j = X_{j'}$ for $j \neq j'$, then the previous numbers are equal. Both these facts give by definitions the identity in (9.14). On the other hand, we always have

$$\mathscr{F}(h) \leqslant n^{-1}\sum_{j=1}^{n} G^{(k)}(X_j/h). \tag{9.16}$$

Since $\|G^{(k)}\|_\infty \leqslant 2$, we conclude by (9.16) that $\mathscr{F}(h) \leqslant 2$. Suppose now that, for some $h_0 > 0$, $\mathscr{F}(h_0) = 0$. This implies that $\omega(f_{n,h_0}; h_0) = 0$, whence f_{n,h_0} should be a constant function with bounded support but this is impossible since f_{n,h_0} is a density. To prove (9.15) consider the function

$$F(x_1, \ldots, x_n) = n^{-1} \sum_{j=1}^{n} G^{(k)}(x_j).$$

It is periodic in each variable, and therefore by a theorem of S. Mazur and W. Orlicz (see [10], p. 15) the equality (9.15) follows.

10. Comments. The procedure of finding the optimal size of the window parameter as described in the previous section has been implemented by L. Chańko, T. Figiel and the author on the IBM PC, and the ESTIMPACK package of programs is available.

REFERENCES

[1] H. Berens and G. Lorentz, *Inverse theorems for Bernstein polynomials*, Indiana Univ. Math. J. 21 (1973), pp. 693–708.

[2] C. de Boor, *Linear combinations of B-splines*, pp. 1–47 in: *Approximation Theory II*, Academic Press, 1976.

[3] Z. Ciesielski, *Constructive function theory and spline systems*, Studia Math. 53 (1975), pp. 277–302.

[4] — *Spline bases in spaces of analytic functions*, pp. 81–111 in: *Canadian Mathematical Society Conference Proceedings*, Vol. 3 (1983): *Second Edmonton Conference on Approximation Theory*, June 7–11, 1982.

[5] — *Local spline approximation and nonparametric density estimation*, pp. 79–84 in: *Constructive Theory of Functions*, Sofia 1987.

[6] — *Nonparametric polynomial density estimation in the L^p norm*, IM PAN, preprint 388 (1987), pp. 1–10. *Proceedings of the Seminar on Approximation and Optimization*, Havana 1987: 01.12.–01.16; LNM 1354, Springer-Verlag, 1988.

[7] — *Nonparametric polynomial density estimation*, Probab. Math. Statist. 9 (1988), pp. 1–10.

[8] — *Haar system and nonparametric density estimation in several variables*, ibidem 9 (1988), pp. 1–11.

[9] L. Devroye and L. Györfi, *Nonparametric Density Estimation. The L_1 View*, J. Wiley, 1985.

[10] S. Mazur et W. O. Orlicz, *Sur quelques propriétés de fonctions périodiques et presque-périodiques*, Studia Math. 9 (1940), pp. 1–16.

[11] E. A. Nadaraja, *Nonparametric estimation of density and of regression curve*, Tbilisi 1983 (in Russian).

[12] I. J. Schoenberg, *Cardinal interpolation and spline function*, J. Approx. Theory 2 (1969), pp. 167–206.

[13] — *Cardinal spline interpolation*, Regional Conference Series 12, SIAM 1973.

[14] L. L. Schumaker, *Spline Functions: Basic Theory*, J. Wiley, 1981.

[15] B. W. Silverman, *Density Estimation for Statistics and Data Analysis*, Chapman and Hall, 1986.

Instytut Matematyczny PAN
Oddział w Gdańsku
ul. Abrahama 18
81-825 Sopot, Poland

Received on 13.6.1989