

T. RYCHLIK (Sopot)

## A CLASS OF UNBIASED KERNEL ESTIMATES OF A PROBABILITY DENSITY FUNCTION

*Abstract.* We propose a class of unbiased and strongly consistent non-parametric kernel estimates of a probability density function, based on a random choice of the sample size and the kernel function. The expected sample size can be arbitrarily small and mild conditions on the local behavior of the density function are imposed.

**1. Introduction and notation.** Unbiased (and minimum variance) estimates of probability density functions have been determined for various parametric models. Some general methods and a number of examples were presented in Voinov and Nikulin [19]. For a rather peculiar nonparametric family of density functions Devroye [4] also determined an unbiased estimate. However, in a general nonparametric setup we have some negative results. Rosenblatt [13], in the fundamental paper where kernel density estimates were introduced, proved that for no sample size does there exist an estimate which is unbiased for the values of all continuous density functions at a given point. Bickel and Lehmann [2] established the same for any family of density functions containing the convex combinations of doubly exponential densities with the scale parameter. Yamato [20] showed that no kernel estimate is unbiased for any density function.

We propose a global unbiased estimate for a general nonparametric class of densities, applying a random rule of choosing a number of observations as well as an element of a sequence of kernel estimates. The rule is independent of the sample. A similar approach was employed in Rychlik [14] to estimate the derivative of a regression function.

We consider a sequence  $X_n$ ,  $n \geq 1$ , of independent real-valued random variables with a common probability density function  $f$ . In order to

---

1991 *Mathematics Subject Classification*: Primary 62G07; Secondary 62L12.

*Key words and phrases*: probability density function, nonparametric estimate, unbiased estimate, randomized estimate, kernel function, rectangular kernel.

construct the estimate we introduce the following notions. We take three sequences of *strictly positive* numbers:

$$(1) \quad \begin{aligned} a_n, \quad n \geq 0, \quad &\text{decreasing to zero,} \\ k_n, \quad n \geq 1, \quad &\text{where all } k_n \text{ are integers,} \\ p_n, \quad n \geq 1, \quad &\text{such that } \sum_{n=1}^{\infty} p_n = 1. \end{aligned}$$

Let  $K$  denote a kernel function which is subject to the conditions:

$$(2) \quad \int K(x) dx = 1, \quad \int K(x)^2 dx = A^2 < \infty, \\ K \text{ is symmetric about the origin, nonincreasing on } [0, \infty), \\ \text{and has a bounded support.}$$

For simplicity we assume that  $\text{supp } K = [-1, 1]$ . With the notation

$$K_n(t) = \frac{1}{a_n} K\left(\frac{t}{a_n}\right), \quad n \geq 0,$$

we define a sequence of kernels by

$$(3) \quad L_n = K_0 + \frac{K_n - K_{n-1}}{p_n}, \quad n \geq 1,$$

and a sequence of estimates

$$\hat{f}_n(x) = \hat{f}_n(X_1, \dots, X_{k_n}; x) = \frac{1}{k_n} \sum_{i=1}^{k_n} L_n(x - X_i), \quad n \geq 1.$$

We see that the  $a_n$  are the bandwidths of the kernel estimates  $K_n$ ,  $n \geq 0$ , and  $k_n$  are the numbers of observations necessary to determine the respective estimates  $\hat{f}_n$ ,  $n \geq 1$ . The sequence  $p_n$ ,  $n \geq 1$ , is used to define the modified kernels in (3), and also to determine the distribution of a positive integer random variable  $N$ , independent of the sequence  $X_n$ ,  $n \geq 1$ ,

$$P(N = n) = p_n, \quad n \geq 1.$$

A randomly chosen  $\hat{f}_N$ , by means of an independent selection rule  $N$ , is our candidate for the unbiased estimate of the density function. Observe that  $\hat{f}_N$  is a global estimate and it integrates to 1, as does each of  $K_n$ ,  $L_n$ , and  $f_n$ . However, this is not a proper probability density function, because it takes negative values. Negative-valued estimates of density functions, very popular nowadays, were introduced in order to reduce the bias and square error and to estimate the derivatives of densities (see, e.g., Bartlett [1], Schucany and Sommers [15], Schuster [16], Silverman [17]). They appear naturally in estimating by means of orthogonal series (see, e.g., Chentsov [3], and Kronmal and Tarter [10]). Koronacki [9] reduced the mean square error by

introducing a kernel estimate which was both negative-valued and randomized. Gajek [8] proposed projecting estimates on the class of proper density functions. The method improves some properties of estimates, but it makes our estimate biased.

In the sequel, we need some auxiliary notation. Let

$$\begin{aligned}
 b_n &= 1 - \frac{a_n}{a_{n-1}}, \\
 c_n &= \frac{1}{\sqrt{k_n \min\{a_{n-1} - a_n, a_n\}}}, \\
 q_n(x) &= EK_n(x - X_1), \\
 r_n(x) &= E\widehat{f}_n(x) = EL_n(x - X_1), \quad n \geq 1.
 \end{aligned}
 \tag{4}$$

For  $f_s(x; t) = \frac{1}{2}[f(x + t) + f(x - t)]$ , the symmetrized density function, we define the oscillation around a point  $x$  by

$$\omega(h) = \omega(f_s, x; h) = \sup_{0 \leq t \leq h} |f_s(x; t) - f(x)|, \quad h > 0.$$

To illustrate our idea of estimation, we focus attention on the simple rectangular kernel  $K = \frac{1}{2}I_{[-1,1]}$ . If  $E\widehat{f}_N(x)$  were finite, we would write

$$\begin{aligned}
 E\widehat{f}_N(x) &= \sum_{n=1}^{\infty} r_n(x)p_n \\
 &= \sum_{n=1}^{\infty} [q_0(x)p_n + q_n(x) - q_{n-1}(x)] \\
 &= \lim_{n \rightarrow \infty} q_n(x) = \lim_{n \rightarrow \infty} \frac{P_n(x)}{2a_n} = f(x)
 \end{aligned}
 \tag{5}$$

for every Lebesgue point  $x$  of  $f$ , with  $P_n(x)$  denoting the probability that  $x - a_n \leq X_1 \leq x + a_n$ . The same idea, with a slightly redefined estimate, carries over naturally to the multidimensional case, but we will not develop this point here.

Our main purpose is to prove that  $E\widehat{f}_N(x)$  exists, and, in consequence, equals  $f(x)$ , under mildest possible smoothness conditions on  $f$ , and for a smallest possible average number of observations. The proof falls naturally into two parts. One consists in evaluating  $E|r_N(x)|$ , which we call the deterministic term, because it depends merely on the choice of parameters of the estimate and local properties of the density function. In the other part, where the stochastic term  $E|\widehat{f}_N(x) - r_N(x)|$  is handled, we also analyze the behavior of observations. In Section 2, assumptions ensuring the finiteness of the deterministic and stochastic terms are stated in Propositions 1 and 2, respectively, and discussed. In particular, Proposition 4 shows that for the naive rectangular kernel  $K$  the assumptions of Proposition 2 are practically

necessary. Our main result is formulated in Proposition 3. In Corollary 1 we notice that a natural modification of our estimate yields additionally the strong consistency. All proofs are given in Section 3.

## 2. Results and discussion

PROPOSITION 1. *If there is an integer  $m = m(x)$  such that*

$$(6) \quad \sum_{n>m} b_n \omega(f_s, x; a_{n-1}) < \infty,$$

then  $E|r_N(x)| < \infty$ .

PROPOSITION 2. *If  $\omega(f_s, x; h) \rightarrow 0$ , as  $h \rightarrow 0$ , and*

$$(7) \quad \sum_{n=1}^{\infty} b_n c_n < \infty,$$

then  $E|\hat{f}_N(x) - r_N(x)| < \infty$ .

Observe that  $a_n = a_0 \prod_{i=1}^n (1 - b_i)$ ,  $n \geq 1$ , tends to zero iff

$$(8) \quad \sum_{n=1}^{\infty} b_n = \infty.$$

Therefore both  $\omega(a_{n-1})$  and  $c_n$ ,  $n \geq 1$ , should decrease sufficiently fast to ensure the convergence of the series (6) and (7). In particular,  $\omega(a_n)$ ,  $n \geq 0$ , is monotone, and so (6) implies the first assumption of Proposition 2, and, in consequence, we have

PROPOSITION 3. *If (7) holds, then  $E\hat{f}_N(x) = f(x)$  for all  $x$  satisfying (6) for some  $m(x) \geq 1$ .*

We now discuss our assumptions in more detail.

If we took into account asymmetric kernels, we would deal with the oscillation of the original density function. Taking symmetric ones, we can confine ourselves to the symmetrized version  $f_s(x; t)$  of the density function. The original  $f$  may be discontinuous at the point under study, whereas  $f_s(x; t)$  should satisfy (6), which is slightly stronger than continuity at zero. Generally,  $\omega(f_s, x; h) \leq \omega(f, x; h)$ , and equality holds for functions  $f$  (locally) symmetric about  $x$ , while  $\omega(f_s, x; h) = 0$  for (locally) asymmetric, e.g. linear, ones. Therefore, at points of smoothness we can estimate better than at sharp local extremes.

In Section 3 we show that if  $\liminf_{n \rightarrow \infty} a_n/a_{n-1} > 0$ , then (6) holds iff

$$(9) \quad \int_0^h \frac{\omega(t)}{t} dt < \infty \quad \text{for some } h > 0.$$

Consequently, if  $a_n, n \geq 0$ , decreases geometrically or slower, then  $\hat{f}_N$  is unbiased at every point  $x$  where

$$\omega(f_s, x; h) = \mathcal{O}\left(\ln^{-1} \frac{1}{h} \dots \ln_k^{-1-\varepsilon} \frac{1}{h}\right) \quad \text{as } h \rightarrow 0,$$

for all  $k \geq 1$  and  $\varepsilon > 0$ . (We adopt the convention that  $\ln_0 x = x_+$  and  $\ln_{k+1} x = \ln(\ln_k x)_+$ .) This is satisfied, with a possible exception of several points, by the density functions occurring in practice. The advantage of condition (9) lies also in the fact of being independent of the parameters of the estimate.

Theoretically, we can extend the set of points with the property of unbiasedness by taking bandwidths, which decrease faster. E.g., choosing  $a_n = a^{-n^2}, n \geq 0, a > 1$ , we also include  $x$ 's such that

$$\omega(f_s, x; h) = \mathcal{O}\left(\ln^{-1/2} \frac{1}{h} \ln_2^{-1} \frac{1}{h} \dots \ln_k^{-1-\varepsilon} \frac{1}{h}\right) \quad \text{as } h \rightarrow 0.$$

Other ways of relaxing conditions on  $f$  consist in modifying assumptions, which would also need reformulating conditions on  $K$  and different reasoning. By analogy, we refer to various proofs of approximating the identity by integral kernel operators (cf. Parzen [12], Stein [18], and Devroye and Wagner [6]). In our proof, no advanced tools of the theory of differentiability of integrals are used. Accordingly, we do not present abstract conclusions, which hold true almost everywhere, but we describe explicitly the points with the desired properties.

For rectangular kernels, (7) becomes necessary if we exclude oscillatory sequences  $c_n, n \geq 1$ , which are unbounded and approach zero.

PROPOSITION 4. *If  $K = \frac{1}{2}I_{[-1,1]}$  and either  $\limsup_{n \rightarrow \infty} c_n < \infty$  or  $\liminf_{n \rightarrow \infty} c_n > 0$ , then  $\sum_{n=1}^{\infty} b_n c_n = \infty$  implies*

$$E|\hat{f}_N(x) - r_N(x)| = \infty$$

for all  $x$  such that  $f(x) > 0$  and  $\omega(f_s, x; h) \rightarrow 0$  as  $h \rightarrow 0$ .

The choice of rectangulars, though natural and easily tractable, is far from optimality in our case. In fact, the rectangulars are the extreme points of the convex class of kernels (2) and so maximize the convex functional  $K \mapsto E|\hat{f}_N(x) - r_N(x)|$  (with  $a_n, k_n, p_n, n \geq 1$ , fixed). Some indirect arguments show that the kernels concentrated about the origin are more advisable. For instance, the major contribution to the summands of the stochastic term with large indices comes from the values of  $\int |K_n - K_{n-1}|$ , which increase with  $\beta$  for power kernels  $K(x) = c_\beta(1 - |x|^\beta)_+, -1/2 < \beta \neq 0$ .

Observe that the integral also becomes smaller as  $a_n/a_{n-1}$  approaches 1. Yet another argument for taking slowly decreasing bandwidths is as follows. Though we are able to choose  $k_n, n \geq 1$ , so that (7) holds for any sequence

$a_n, n \geq 0$ , we are interested in minimizing the sample size. If  $a_n, n \geq 0$ , vanishes slower, we can preserve (7) by taking slower increasing  $k_n, n \geq 1$  (we deduce from (1) and (8) that  $\limsup_{n \rightarrow \infty} k_n = \infty$  is necessary). For instance, for  $a_n = a^{-n}, n^{-\alpha}$  and  $\ln_j^{-1} n$ , where  $a > 1$  and  $\alpha > 0$ , it suffices to take  $k_n = \mathcal{O}(a^{n/2} n^2 \ln^2 n \dots \ln_k^{2+\varepsilon} n)$ ,  $\mathcal{O}(n^{1+\alpha} \ln^2 n \dots \ln_k^{2+\varepsilon} n)$  and  $\mathcal{O}(n \ln n \dots \ln_{j-1} n \ln_j^2 n \dots \ln_k^{2+\varepsilon} n)$ , respectively. Putting  $\varepsilon = 0$  in the above expressions gives divergence of the series and infiniteness of the stochastic term.

An important point is that Proposition 3 does not involve any additional assumptions about the distribution of the stopping rule  $N$ . All  $p_n, n \geq 1$ , are merely required to be positive. We cannot avoid obtaining large  $N$  and large  $k_N$ , but we can make their probabilities arbitrarily small. Also, it is even possible to make  $Ek_N$  as close to 1 as we wish. Theoretically, we should have a possibility of performing arbitrarily many independent experiments, but choosing a stopping rule properly, we can make it almost unlikely to have more than very few repetitions.

Summarizing the above considerations, we recommend the randomized estimate based on  $\hat{f}_n, n \geq 1$ , with bandwidths slowly approaching zero, and moderately increasing sample sizes, a kernel mostly concentrated about the origin, and a stopping rule with an overwhelming probability mass attached to several first elements.

Finally, we point out that our Proposition 3, together with the strong law of large numbers, suggests a standard construction of a strongly consistent unbiased estimate of density functions. Let  $N_j, j \geq 1$ , be independent random variables distributed as  $N$ ,  $S_0 = 0$ , and  $S_k = \sum_{j=1}^k N_j, k \geq 1$ .

COROLLARY 1. *Under the hypotheses of Proposition 3,*

$$\frac{1}{k} \sum_{j=1}^k \hat{f}_{N_j}(X_{S_{j-1}+1}, \dots, X_{S_j}; x) \rightarrow f(x) \quad \text{as } k \rightarrow \infty,$$

*with probability one.*

One may be interested in the rate of convergence, which is intimately related to existence of higher moments  $E|\hat{f}_N(x) - r_N(x)|^p$  for  $p > 1$ . This is possible to be accomplished by taking more observations so that each  $\hat{f}_n, n \geq 1$ , approximates the density function more precisely. Since, moreover, all moments except the first one depend on the distribution of  $N$ , the sample size  $k_N$  is expected to increase dramatically.

One may conjecture that the estimation error will be reduced by using another stopping rule. In our approach, the sample size is determined independently of observations. A properly constructed sequential rule makes use of the information carried by each consecutive experiment and stops

sampling when an estimate of error becomes sufficiently small. However, preservation of unbiasedness and reduction of estimation errors are, in some sense, contradictory requirements (cf., e.g., Doss and Sethuraman [7], and Liu and Brown [11]).

**3. Proofs.** Proof of Proposition 1. Let a fixed  $x$  satisfy (6) and  $m$  be the smallest integer such that  $\omega(a_m)$  is finite. Then

$$\begin{aligned} (10) \quad q_n(x) &= \int K_n(t) f_s(x; t) dt \\ &\leq [f(x) + \omega(a_m)] \int_{|t| \leq a_m} K_n(t) dt + K_n(a_m) \int_{|t| > a_m} f_s(x; t) dt \\ &\leq f(x) + \omega(a_m) + K_n(a_m) < \infty \end{aligned}$$

for all  $n \geq 1$ . Also,

$$\begin{aligned} q_n(x) - q_{n-1}(x) &= \int [K_n(t) - K_{n-1}(t)] [f_s(x; t) - f(x)] dt \\ &= b_n \int \left[ K_n(t) + \frac{a_n K_n(t) - a_{n-1} K_{n-1}(t)}{a_{n-1} - a_n} \right] [f_s(x; t) - f(x)] dt. \end{aligned}$$

Since  $K_n$  is nonnegative and the fraction is nonpositive, and both vanish for  $|t| > a_{n-1}$ , we obtain

$$\begin{aligned} (11) \quad |q_n(x) - q_{n-1}(x)| &\leq b_n \omega(a_{n-1}) \int \left[ K_n(t) + \frac{a_{n-1} K_{n-1}(t) - a_n K_n(t)}{a_{n-1} - a_n} \right] dt \\ &= 2b_n \omega(a_{n-1}). \end{aligned}$$

Therefore, by (6), (10), and (11),

$$\begin{aligned} E|r_N(x)| &= \sum_{n=1}^{\infty} \left| q_0(x) + \frac{q_n(x) - q_{n-1}(x)}{p_n} \right| p_n \\ &\leq q_0(x) + \sum_{n=1}^{\infty} |q_n(x) - q_{n-1}(x)| \\ &\leq q_0(x) + \sum_{n=1}^m [q_n(x) + q_{n-1}(x)] + 2 \sum_{n>m} b_n \omega(a_{n-1}) < \infty. \blacksquare \end{aligned}$$

**Proof of Proposition 2.** We first examine absolute deviations of  $\hat{f}_n(x)$ ,  $n \geq 1$ . Generally, by the Jensen inequality for the reverse martingale of means of independent identically distributed random variables, we obtain

$$\begin{aligned} (12) \quad E|\hat{f}_n(x) - r_n(x)| &\leq E|L_n(x - X_1) - r_n(x)| \\ &\leq 2 \left[ q_0(x) + \frac{q_n(x) + q_{n-1}(x)}{p_n} \right] < \infty. \end{aligned}$$

For large  $n$ , we need a refined evaluation. Since

$$L_n = K_0 - \frac{1}{p_n} \left( K_{n-1} - \frac{a_n}{a_{n-1}} K_n \right) + \frac{b_n K_n}{p_n},$$

we have

$$\begin{aligned} (13) \quad E|\widehat{f}_n(x) - r_n(x)| &\leq \frac{1}{k_n} E \left| \sum_{i=1}^{k_n} [K_0(x - X_i) - q_0(x)] \right| \\ &\quad + \frac{1}{k_n a_{n-1} p_n} E \left| \sum_{i=1}^{k_n} [a_{n-1} K_{n-1}(x - X_i) \right. \\ &\quad \left. - a_n K_n(x - X_i) - a_{n-1} q_{n-1}(x) + a_n q_n(x)] \right| \\ &\quad + \frac{b_n}{k_n p_n} E \left| \sum_{i=1}^{k_n} K_n(x - X_i) - q_n(x) \right| \\ &= I_1 + I_2 + I_3 \quad (\text{say}). \end{aligned}$$

Below, we estimate each summand separately. First, repeating the arguments of (12), we obtain

$$(14) \quad I_1 \leq 2q_0(x).$$

By the Schwarz inequality and independence of observations,

$$\begin{aligned} (k_n a_{n-1} p_n I_2)^2 &\leq k_n \operatorname{Var}[a_{n-1} K_{n-1}(x - X_1) - a_n K_n(x - X_1)] \\ &\leq k_n E[a_{n-1} K_{n-1}(x - X_1) - a_n K_n(x - X_1)]^2. \end{aligned}$$

Applying monotonicity of the kernel function and boundedness of its support and (2), we deduce that

$$\begin{aligned} (15) \quad k_n (a_{n-1} p_n I_2)^2 &\leq E \left[ K^2 \left( \frac{x - X_1}{a_{n-1}} \right) - K^2 \left( \frac{x - X_1}{a_n} \right) \right] \\ &\leq [f(x) + \omega(a_{n-1})] \int \left[ K^2 \left( \frac{t}{a_{n-1}} \right) - K^2 \left( \frac{t}{a_n} \right) \right] dt \\ &= [f(x) + \omega(a_{n-1})] A^2 (a_{n-1} - a_n). \end{aligned}$$

Similarly,

$$(16) \quad k_n \left( \frac{p_n I_3}{b_n} \right)^2 \leq E K_n(x - X_1)^2 \leq \frac{[f(x) + \omega(a_n)] A^2}{a_n}.$$

Combining (14)–(16) with (4), we can rewrite (13) as

$$E|\widehat{f}_n(x) - r_n(x)| \leq 2q_0(x) + 2\sqrt{f(x) + \omega(a_{n-1})} A \frac{b_n c_n}{p_n}.$$



Finally, we see that

$$\begin{aligned} E|\widehat{f}_N(x) - r_N(x)| &= \sum_{n=1}^{\infty} E|\widehat{f}_n(x) - r_n(x)|p_n \\ &\leq 2q_0(x) + 2\sum_{n=1}^m [q_n(x) + q_{n-1}(x)] \\ &\quad + 2\sqrt{f(x) + \omega(a_m)}A \sum_{n>m} b_n c_n < \infty. \blacksquare \end{aligned}$$

**Proof of Proposition 3.** By Propositions 1 and 2,  $E\widehat{f}_N(x)$  is well defined, and satisfies

$$E\widehat{f}_N(x) = \lim_{n \rightarrow \infty} q_n(x)$$

(cf. (5)). This is exactly  $f(x)$ , because

$$|q_n(x) - f(x)| = \left| \int K_n(t)[f_s(x;t) - f(x)] dt \right| \leq \omega(a_n) \rightarrow 0$$

as  $n \rightarrow \infty$ , and the proposition follows.  $\blacksquare$

**Proof of (6) $\Leftrightarrow$ (9) for  $\liminf a_n/a_{n-1} > 0$ .** Take a piecewise linear function  $\alpha$  such that  $\alpha(1/n) = a_n$  and  $\alpha'(t) = n(n-1)(a_{n-1} - a_n)$  for  $t \in (1/(n-1), 1/n)$ . By monotonicity of  $\alpha$  and  $\omega$ , changing variables we obtain

$$\begin{aligned} \int_{a_n}^{a_{n-1}} \frac{\omega(t)}{t} dt &= c \int_{1/n}^{1/(n-1)} \frac{\alpha'(s)}{\alpha(s)} \omega(\alpha(s)) ds \\ &\leq \int_{1/n}^{1/(n-1)} \frac{\alpha'(s)}{a_{n-1}} \omega(a_{n-1}) ds = b_n \omega(a_{n-1}) \\ &\leq \int_{1/n}^{1/(n-1)} \frac{\alpha'(s)}{\alpha(s)} \omega\left(\frac{\alpha(s)}{c}\right) ds = \int_{a_n/c}^{a_{n-1}/c} \frac{\omega(t)}{t} dt \end{aligned}$$

for all sufficiently large  $n$  (say  $n > m$ ), and a sufficiently small  $c$  such that  $ca_{n-1} \leq a_n$  for every  $n > m$ . Consequently,

$$c \int_0^{a_m} \frac{\omega(t)}{t} dt \leq \sum_{n>m} b_n \omega(a_{n-1}) \leq \int_0^{a_m/c} \frac{\omega(t)}{t} dt,$$

which is our assertion.  $\blacksquare$

**Proof of Proposition 4.** Write

$$\widehat{f}_n(x) - r_n(x) = g_n(x) + \frac{h_n(x)}{p_n},$$

where

$$g_n(x) = \frac{1}{k_n} \sum_{i=1}^{k_n} K_0(x - X_i) - q_0(x),$$

and

$$(17) \quad h_n(x) = \frac{1}{k_n} \sum_{i=1}^{k_n} [K_n(x - X_i) - K_{n-1}(x - X_i)] - q_n(x) + q_{n-1}(x).$$

By (14),  $E|g_n(x)| = I_1 < \infty$ ,  $n \geq 1$ , and  $E|g_N(x)| < \infty$ . Since

$$E|\hat{f}_N(x) - r_N(x)| \geq E \frac{|h_N(x)|}{p_N} - E|g_N(x)|,$$

we need to show that

$$E \frac{|h_N(x)|}{p_N} = \sum_{n=1}^{\infty} |h_n(x)| = \infty.$$

To this end, we apply the following lower bound for the first central absolute moment of the sample mean.

LEMMA 1. *Let  $Z_1, \dots, Z_k$  be independent identically distributed random variables such that  $EZ_1 = 0$  and  $P(|Z_1 - Z_2| \geq a) \geq p$ . If  $kp \geq c > 0$ , then*

$$E \left| \frac{1}{k} \sum_{i=1}^k Z_i \right| \geq \frac{ac}{4(4+c)} \sqrt{\frac{p}{k}}.$$

This is a modification of an inequality presented by Devroye and Györfi [5, Chapter 5, Lemma 27]. We shall not prove it here, because actually it is an essential part of the proof in [5].

For the rectangular kernel  $K = \frac{1}{2}I_{[-1,1]}$ , we introduce a sequence of symmetric random variables

$$S_n(x) = K_n(x - X_1) - K_{n-1}(x - X_1) - K_n(x - X_2) + K_{n-1}(x - X_2),$$

$n \geq 1$ , whose probabilities of positive values are

$$P\left(S_n(x) = \frac{1}{2a_n}\right) = P_n(x)[P_{n-1}(x) - P_n(x)],$$

$$P\left(S_n(x) = \frac{1}{2a_{n-1}}\right) = [P_{n-1}(x) - P_n(x)][1 - P_{n-1}(x)],$$

$$P\left(S_n(x) = \frac{1}{2a_n} - \frac{1}{2a_{n-1}}\right) = P_n(x)[1 - P_{n-1}(x)].$$

If the last two values coincide, we obviously sum up the respective probabilities.

Assume first that  $\limsup_{n \rightarrow \infty} c_n < \infty$ , and take  $m \geq 1$  such that  $\omega(a_m) < f(x)$ , and  $c_n \leq C$  for some  $C > 0$  and all  $n > m$ .

If  $a_{n-1} \leq 2a_n$ , then

$$\begin{aligned} P\left(|S_n(x)| \geq \frac{1}{2a_{n-1}}\right) &= P\left(|S_n(x)| = \frac{1}{2a_{n-1}} \text{ or } \frac{1}{2a_n}\right) \\ &\geq 2[P_{n-1}(x) - P_n(x)][1 - P_{n-1}(x) + P_n(x)] \\ &\geq 4[f(x) - \omega(a_m)][1 - P_m(x)](a_{n-1} - a_n) \\ &= 4B(a_{n-1} - a_n) \quad (\text{say}), \end{aligned}$$

and

$$4Bk_n(a_{n-1} - a_n) = \frac{4B}{c_n^2} \geq \frac{4B}{C^2}.$$

By Lemma 1,

$$(18) \quad E|h_n(x)| \geq \frac{B^{3/2}}{4(B + C^2)} \cdot \frac{b_n}{\sqrt{(a_{n-1} - a_n)k_n}} = \frac{B^{3/2}}{4(B + C^2)} b_n c_n.$$

Likewise, for  $a_{n-1} > 2a_n$  we have

$$\begin{aligned} P\left(|S_n(x)| \geq \frac{1}{2a_n} - \frac{1}{2a_{n-1}}\right) &= P\left(|S_n(x)| = \frac{1}{2a_n} - \frac{1}{2a_{n-1}} \text{ or } \frac{1}{2a_n}\right) \\ &= 2P_n(x)[1 - P_n(x)] \geq 4Ba_n, \end{aligned}$$

and

$$4Bk_n a_n = \frac{4B}{c_n^2} \geq \frac{4B}{C^2}.$$

Applying Lemma 1 again, we obtain

$$E|h_n(x)| \geq \frac{B^{3/2}}{4(B + C^2)} \cdot \frac{b_n}{\sqrt{a_n k_n}} = \frac{B^{3/2}}{4(B + C^2)} b_n c_n.$$

This, together with (18), yields

$$\sum_{n=1}^{\infty} E|h_n(x)| \geq \frac{B^{3/2}}{4(B + C^2)} \sum_{n>m} b_n c_n = \infty,$$

which is the desired assertion.

Suppose now that  $\liminf_{n \rightarrow \infty} c_n > 0$ , i.e., there exists a positive  $c$  such that  $c_n \geq c$  for all  $n > m$ . Then we take  $k'_n \geq k_n$ ,  $n > m$ , such that

$$\frac{c}{2} \leq c'_n = \frac{1}{\sqrt{\min\{a_{n-1} - a_n, a_n\}k'_n}} \leq c,$$

and define  $h'_n(x)$  by writing  $k'_n$  instead of  $k_n$  in (17). Applying the Jensen inequality for means and the arguments of the proof for the previous case, with  $k_n$  replaced by  $k'_n$ , we conclude that

$$E|h_n(x)| \geq E|h'_n(x)| \geq \frac{B^{3/2}}{4(B + c^2)} b_n c'_n \geq \frac{B^{3/2}c}{8(B + c^2)} b_n.$$

Finally, by (8),

$$\sum_{n=1}^{\infty} |h_n(x)| \geq \frac{B^{3/2}c}{8(B+c^2)} \sum_{n>m} b_n = \infty,$$

and the proof of Proposition 4 is complete. ■

**Acknowledgements.** The author thanks the referee for drawing his attention to references [8] and [9].

### References

- [1] M. S. Bartlett, *Statistical estimation of density functions*, Sankhyā Ser. A 25 (1963), 245–254.
- [2] P. Bickel and E. Lehmann, *Unbiased estimation in convex families*, Ann. Math. Statist. 40 (1969), 1523–1535.
- [3] N. N. Chentsov, *An estimate of an unknown probability density under observations*, Dokl. Akad. Nauk SSSR 147 (1962), 45–48 (in Russian).
- [4] L. P. Devroye, *A Course in Density Estimation*, Birkhäuser, Boston, 1987.
- [5] L. P. Devroye and L. Györfi, *Nonparametric Density Estimation. The  $L_1$  View*, Wiley, New York, 1985.
- [6] L. P. Devroye and T. J. Wagner, *The  $L_1$  convergence of kernel density estimates*, Ann. Statist. 7 (1979), 1136–1139.
- [7] H. Doss and J. Sethuraman, *The price of bias reduction when there is no unbiased estimate*, *ibid.* 17 (1989), 440–442.
- [8] L. Gajek, *On improving density estimators which are not bona fide functions*, *ibid.* 14 (1986), 1612–1618.
- [9] J. Koronacki, *Kernel estimation of smooth densities using Fabian’s approach*, Statistics 18 (1987), 37–47.
- [10] R. Kronmal and M. Tarter, *The estimation of probability densities and cumulatives by Fourier series methods*, J. Amer. Statist. Assoc. 63 (1968), 925–952.
- [11] R. C. Liu and L. D. Brown, *Nonexistence of informative unbiased estimators in singular problems*, Ann. Statist. 21 (1993), 1–13.
- [12] E. Parzen, *On estimation of a probability density function and mode*, Ann. Math. Statist. 33 (1962), 1065–1076.
- [13] M. Rosenblatt, *Remarks on some nonparametric estimates of a density function*, *ibid.* 27 (1956), 832–837.
- [14] T. Rychlik, *Unbiased nonparametric estimation of the derivative of the mean*, Statist. Probab. Lett. 10 (1990), 329–333.
- [15] W. R. Schucany and J. P. Sommers, *Improvement of kernel type density estimators*, J. Amer. Statist. Assoc. 72 (1977), 420–423.
- [16] E. F. Schuster, *Estimation of a probability density function and its derivatives*, Ann. Math. Statist. 40 (1969), 1187–1195.
- [17] B. W. Silverman, *Weak and strong uniform consistency of the kernel estimate of a density and its derivatives*, Ann. Statist. 6 (1978), 177–184.
- [18] E. M. Stein, *Singular Integrals and Differentiability Properties of Functions*, Princeton Univ. Press, Princeton, N.J., 1970.
- [19] V. G. Voinov and M. S. Nikulin, *Unbiased Estimators and their Applications*, Vol. 1, *Univariate Case*, Kluwer Academic Publ., Dordrecht, 1993.

- [20] H. Yamato, *Some statistical properties of estimators of density and distribution functions*, Bull. Math. Statist. 15 (1972), 113–131.

TOMASZ RYCHLIK  
INSTITUTE OF MATHEMATICS  
POLISH ACADEMY OF SCIENCES  
ABRAHAMA 18  
81-825 SOPOT, POLAND

*Received on 10.1.1994*