

KAROL DZIEDZIUL and PIOTR PALUSZEK (Gdańsk)

BOX-SPLINE HISTOGRAMS FOR MULTIVARIATE DENSITY ESTIMATION

Abstract. The uniform approach to calculation of MISE for histogram and density box-spline estimators gives us a possibility to obtain estimators of derivatives of densities and the asymptotic constant.

1. Histograms. We have two methods of bandwidth selection for the Rosenblatt–Parzen estimator: cross-validation (unbiased and biased) and plug-in (see for instance [16]). In our paper we present higher order point estimators to obtain plug-in estimates for the bandwidth in the case of histograms. An excellent introduction to histograms is given in the book by Scott [17].

In Section 1 we give a simple introduction to box-spline operators and box-spline histograms based on these operators. From the point of view of estimation of density, two properties of approximation are crucial: the rate of convergence and the so called saturation property (Theorem 3.1). These properties divide the box-spline estimators into three classes. These classes are represented by a histogram, a linear histogram, and a Zwart–Powell histogram (ZP histogram for short). This is the reason why we fix our attention on these three histograms. The basic results are recalled in Section 2. In Section 3 we show how to use the saturation property to estimate derivatives. The presented method is a version of the method from [13], and it is applicable only to box-spline estimators of the type of the ZP histogram. In Section 4 we present a method of estimating the asymptotic constant (see (11)). This method is more general and is applicable to all cases. We present it for the histogram. It seems to be a version of the “no diagonals” estimator [18].

2010 *Mathematics Subject Classification:* 41A35, 41A60, 62G07, 62G20.

Key words and phrases: histogram, linear histogram, spline-estimator, MISE, bandwidth selection.

We consider only dimension $d = 2$ just for simplicity. It is also not complicated to introduce a nonhomogeneous scaling [9], so we omit it. Spline estimators were introduced by Ciesielski [5]. There are a large number of papers concerning methods of bandwidth selection. Let us mention some of them: [1], [11], [14].

We consider a pair of functions (F, G) which are nonnegative, piecewise polynomials with compact support

$$F, G : \mathbb{R}^2 \rightarrow \mathbb{R}.$$

The additional assumptions are given below. A box-spline operator under consideration is defined by

$$(1) \quad Qf(x) = \int_{\mathbb{R}^2} K(x, y)f(y) dy,$$

with the kernel depending on F, G ,

$$(2) \quad K(x, y) = \sum_{\alpha \in \mathbb{Z}^2} F(y - \alpha)G(x - \alpha).$$

The operator Q defines a family of operators Q_h for $h > 0$,

$$(3) \quad Q_h = \sigma_h \circ Q \circ \sigma_{1/h},$$

where

$$\sigma_h f(x) = f(x_1/h, x_2/h), \quad x = (x_1, x_2).$$

REMARKS. Since $F, G \geq 0$, it follows that the operators Q_h are positive, i.e. if $f \geq 0$ then $Q_h f \geq 0$. Moreover, we assume that our functions satisfy

$$\sum_{\alpha \in \mathbb{Z}^2} F(\cdot - \alpha) = \frac{1}{\int_{\mathbb{R}^2} G} \quad \text{a.e.}$$

By this assumption the operators Q_h map densities to densities, i.e. if f is a density, then $\int_{\mathbb{R}^2} Q_h f = 1$. To ensure a good approximation we assume that the Q_h reproduce at least constant polynomials, $Q_h(1) = 1$ (here 1 is the function $1(x) = 1$), or equivalently

$$\sum_{\alpha \in \mathbb{Z}^2} G(x - \alpha) = \frac{1}{\int_{\mathbb{R}^2} F} \quad \text{a.e.}$$

REMARK. In the definition of the kernel K , instead of \mathbb{Z}^2 we can take another lattice, for instance $A\mathbb{Z}^2$, where

$$(4) \quad A = \begin{bmatrix} 1 & \sin \pi/6 \\ 0 & \cos \pi/6 \end{bmatrix}.$$

The definition so modified includes for instance the histograms and the linear histogram based on the regular hexagon considered in [17].

Let X_1, \dots, X_n be a random sample from a distribution with density f . We define a density estimator based on the kernel K by

$$(5) \quad f_{h,n}(x) = \frac{1}{nh^2} \sum_{k=1}^n K(x/h, X_k/h).$$

Note that

$$(6) \quad Ef_{h,n} = Q_h f.$$

Now we introduce three examples of box-spline estimators. Let H be the characteristic function of the square $[0, 1]^2$, i.e.

$$H(x) = I_{[0,1]^2}(x) = \begin{cases} 1, & x \in [0, 1]^2, \\ 0, & x \notin [0, 1]^2. \end{cases}$$

In this paper we will consider three examples of pairs (F_i, G_i) , $i = 1, 2, 3$.

EXAMPLE 1. *The histogram* corresponds to the choice

$$F_1 = G_1 = H.$$

EXAMPLE 2. *The linear histogram* corresponds to the choice

$$F_2(x_1, x_2) = H(x_1 - 0.5, x_2 - 0.5)$$

and G_2 is the hat function given by

$$G_2(x_1, x_2) = \int_0^1 G_1(x_1 - t, x_2 - t) dt.$$

EXAMPLE 3. *The Zwart–Powell histogram* (for short ZP histogram) corresponds to the choice

$$F_3(x_1, x_2) = H(x_1, x_2 - 1)$$

and G_3 is the Zwart–Powell function given by

$$G_3(x_1, x_2) = \int_0^1 G_2(x_1 + t, x_2 - t) dt.$$

See [3] for the definition of box-splines and [7] for the definition of box-spline estimators.

2. Asymptotic formulas for MISE. We say that the box-spline operator Q reproduces the polynomials of degree less than ϱ if $Q(P) = P$ for all polynomials P with $\deg P < \varrho$. We then say that Q has *polynomial order* ϱ . For a pair of functions (F_j, G_j) introduced in the previous section we will add the superscript j to the operator, i.e. Q^j , and to the kernel i.e. K^j , $j = 1, 2, 3$. Note that the operator Q^1 reproduces only constant polynomials, i.e. $\varrho_1 = 1$, and the operators Q^j , $j = 2, 3$, reproduce the linear and constant polynomials, i.e. $\varrho_j = 2$. The parameter ϱ gives the rate of approximation.



We may check [8] that if Q has polynomial order ϱ , then there is $C > 0$ such that for functions f from the Sobolev space W_2^ϱ ,

$$(7) \quad \|Q_h f - f\|_2 \leq Ch^\varrho |f|_{\varrho,2},$$

where

$$|f|_{\varrho,2} = \sum_{|\beta|=\varrho} \|D^\beta f\|_2, \quad \|f\|_2 = \left(\int_{\mathbb{R}^d} |f|^2 \right)^{1/2}.$$

$$D^\beta f = \frac{\partial^{|\beta|} f}{\partial x_1^{\beta_1} \partial x_2^{\beta_2}}, \quad \beta = (\beta_1, \beta_2), \quad |\beta| = \beta_1 + \beta_2.$$

Recall that the Sobolev space is defined by

$$W_2^\varrho = \left\{ f \in L^2 : \sum_{|\beta|=\varrho} \|D^\beta f\|_2 < \infty \right\}.$$

A monomial of degree $|\beta|$ will be denoted by $[\]^\beta$, i.e. for $x = (x_1, x_2)$,

$$[\]^\beta(x) = x^\beta = x_1^{\beta_1} x_2^{\beta_2}.$$

We assume that $f \in L^2$ to consider the mean integrated square error, given by

$$(8) \quad \text{MISE}(f, h) = E \left[\int_{\mathbb{R}^2} [f_{h,n} - f]^2 \right].$$

Consequently,

$$(9) \quad \text{MISE}(f, h) = E \left[\int_{\mathbb{R}^2} [f_{h,n} - Q_h f]^2 \right] + \int_{\mathbb{R}^2} [Q_h f - f]^2.$$

The deterministic part is considered in [8].

THEOREM 2.1. *Assume that Q has polynomial order ϱ . Let $f \in W_2^\varrho(\mathbb{R}^2)$. Then*

$$(10) \quad \lim_{h \rightarrow 0^+} \left\| \frac{Q_h f - f}{h^\varrho} \right\|_2 = \left(\int_{\mathbb{R}^2} \left(\int_{[0,1]^2} \left| \sum_{|\beta|=\varrho} \frac{1}{\beta!} D^\beta f(t) (Q([\]^\beta)(x) - [\]^\beta(x)) \right|^2 dx \right) dt \right)^{1/2}.$$

Let us define the asymptotic constant depending on f and the box-spline histogram ($j = 1, 2, 3$) by

$$(11) \quad \theta_j = (\text{Asym}(f, j))^2 = \int_{\mathbb{R}^2} \left(\int_{[0,1]^2} \left| \sum_{|\beta|=\varrho_j} \frac{1}{\beta!} D^\beta f(t) (Q^j([\]^\beta)(x) - [\]^\beta(x)) \right|^2 dx \right) dt.$$

Asym($f, 1$) is known (see for instance [17]):

$$(12) \quad \theta_1 = (\text{Asym}(f, 1))^2 = \frac{1}{12} \int_{\mathbb{R}^2} \left(\left(\frac{\partial f}{\partial x_1} \right)^2 + \left(\frac{\partial f}{\partial x_2} \right)^2 \right).$$

An easy calculation shows that

$$(13) \quad \theta_2 = (\text{Asym}(f, 2))^2 = \int_{\mathbb{R}^2} \left\{ \left(\frac{\partial^2 f}{\partial x_1^2} \right)^2 \cdot \frac{49}{2880} + \left(\frac{\partial^2 f}{\partial x_2^2} \right)^2 \cdot \frac{49}{2880} + \left(\frac{\partial^2 f}{\partial x_1 \partial x_2} \right)^2 \cdot \frac{1}{90} + \left(\frac{\partial^2 f}{\partial x_1^2} \frac{\partial^2 f}{\partial x_2^2} \right) \cdot \frac{1}{32} + \left(\frac{\partial^2 f}{\partial x_1^2} \frac{\partial^2 f}{\partial x_1 \partial x_2} \right) \cdot \frac{17}{720} + \left(\frac{\partial^2 f}{\partial x_2^2} \frac{\partial^2 f}{\partial x_1 \partial x_2} \right) \cdot \frac{17}{720} \right\},$$

$$(14) \quad \theta_3 = (\text{Asym}(f, 3))^2 = \int_{\mathbb{R}^2} \left\{ \frac{1}{6} \left(\frac{\partial^2 f}{\partial x_1^2} + \frac{\partial^2 f}{\partial x_2^2} \right) \right\}^2.$$

Let us consider the first term of (9). We have the following result (compare [8, Theorem 1.4]).

THEOREM 2.2. *Let the density f be in $W_2^q(\mathbb{R}^2)$. If $nh^2 \rightarrow \infty$ and $h \rightarrow 0$ then*

$$(15) \quad \lim_{nh^2 \rightarrow \infty} nh^2 E \left[\int_{\mathbb{R}^2} [f_{h,n} - Q_h f]^2 \right] = \int_{\mathbb{R}^2} \left[\int_{[0,1]^2} (K(x, y))^2 dy \right] dx.$$

REMARK 1. By (9), (10) and (15) we get

$$\text{MISE}(f, h) \sim \text{AMISE}(f, h)$$

where for the box-spline histograms

$$(16) \quad \text{AMISE}(f, h) = \frac{1}{nh^2} \int_{\mathbb{R}^2} \left[\int_{[0,1]^2} (K^j(x, y))^2 dy \right] dx + h^{2q_j} (\text{Asym}(f, j))^2.$$

So the best choice of the parameter $h > 0$ to minimize (16) is

$$(17) \quad h = \left(\frac{\int_{\mathbb{R}^2} \left[\int_{[0,1]^2} (K^j(x, y))^2 dy \right] dx}{\varrho_j n (\text{Asym}(f, j))^2} \right)^{-1/(2q_j+2)}.$$

Now we have another estimation problem of $\theta_j = (\text{Asym}(f, j))^2$ with a different bandwidth denoted by a . As for the density estimation, the choice of a is crucial to the performance of the estimator $\hat{\theta}_j(a)$. We use here the notation from [15]. In the next section we construct an estimator g_{an} of the derivatives $D_{Q_3} f$, where (compare (14))

$$D_{Q_3} f = \frac{1}{6} \left(\frac{\partial^2 f}{\partial x_1^2} + \frac{\partial^2 f}{\partial x_2^2} \right).$$

Hence an estimator of θ_3 is given by

$$\widehat{\theta}_3(a) = \int_{\mathbb{R}^2} (g_{an})^2.$$

In Section 4 we estimate θ_j directly.

3. Choice of bandwidth for estimation of derivatives of density: ZP histogram. The problem of estimation of derivatives in the multivariate case is rather ambiguous. We look for estimators of derivatives which appear in the asymptotic formula. The following theorem ([10, Theorem 2.5]) is important. Let us denote $\check{F}(x) = F(-x)$.

THEOREM 3.1. *Let $f \in W_2^{\varrho_j}$. If $a \rightarrow 0$ then*

$$(18) \quad \frac{Q_a^j f - f}{a^{\varrho_j}} \rightarrow D_{Q^j} f$$

weakly in L^2 for $j = 1, 2, 3$, where

$$D_{Q^j} f = \frac{1}{(2\pi i)^{\varrho_j}} \sum_{|\beta|=\varrho_j} \frac{D^\beta f}{\beta!} D^\beta (\widehat{G}_j \widehat{\check{F}}_j)(0).$$

It is crucial for our construction that for $j = 3$ in (18) we have L^2 convergence, but not for $j = 1, 2$ in general. Hence

$$(\text{Asym}(f, 3))^2 = \int_{\mathbb{R}^2} (D_{Q^3} f)^2.$$

It is not difficult to prove (Theorem 3.2 below) that if $a \rightarrow 0$, then also

$$(19) \quad \frac{Q_a^3(Q_a^3 f) - Q_a^3 f}{a^2} \rightarrow D_{Q^3} f = \frac{1}{6} \left(\frac{\partial^2 f}{\partial x_1^2} + \frac{\partial^2 f}{\partial x_2^2} \right),$$

in L^2 norm. The property (19) helps us construct an estimator of $D_{Q^3} f$. It leads us to the operator

$$T := Q^3 \circ Q^3 - Q^3.$$

Note that the operator T has the same structure as Q , i.e.

$$Tf(x) = \int_{\mathbb{R}^2} \kappa(x, y) f(y) dy,$$

where

$$(20) \quad \kappa(x, y) = \sum_{\alpha \in \mathbb{Z}^2} F_3(y - \alpha) (Q^3(G_3) - G_3)(x - \alpha).$$

We define as above a family of operators T_a for $a > 0$. Now we are ready to define an estimator of the derivatives $D_{Q^3} f$. Let X_1, \dots, X_n be a random



sample from a distribution with density f . Then we define an estimator of $D_{Q^3}f$ by

$$(21) \quad g_{an}(x) = \frac{1}{na^4} \sum_{k=1}^n \kappa(x/a, X_k/a).$$

Note that

$$Eg_{an} = \frac{T_a(f)}{a^2}$$

and

$$E \int_{\mathbb{R}^2} |g_{an} - D_{Q^3}f|^2 = E \int_{\mathbb{R}^2} \left| g_{an} - \frac{T_a(f)}{a^2} \right|^2 + \int_{\mathbb{R}^2} \left| \frac{T_a(f)}{a^2} - D_{Q^3}f \right|^2.$$

THEOREM 3.2. *Let $na^6 \rightarrow \infty$ and $a \rightarrow 0$ and $f \in W_2^4$. Then*

$$(22) \quad \left(\int_{\mathbb{R}^2} \left| \frac{T_a(f)}{a^2} - D_{Q^3}f \right|^2 \right)^{1/2} \leq C(a|f|_{3,2} + a^2|f|_{4,2})$$

and

$$\lim_{na^6 \rightarrow \infty} na^6 E \int_{\mathbb{R}^2} \left| g_{an} - \frac{T_a(f)}{a^2} \right|^2 = \int_{\mathbb{R}^2} \left\{ \int_{[0,1]^2} \kappa^2(x, y) dy \right\} dx.$$

Proof. If we modify slightly the end of the proof of Theorem 2.23 with $\varrho = 0$ in [6] and use Lemma 1.1 of [8], we get

$$(23) \quad \left\| \frac{Q_a^3 f - f}{a^2} - D_{Q^3}f \right\|_2 \leq Ca|f|_{3,2}.$$

Now the triangle inequality implies that

$$(24) \quad \begin{aligned} \left\| \frac{T_a(f)}{a^2} - D_{Q^3}f \right\|_2 &= \left\| \frac{Q_a^3 Q_a^3 f - Q_a^3 f}{a^2} - D_{Q^3}f \right\|_2 \\ &\leq \left\| \frac{Q_a^3 Q_a^3 f - Q_a^3 f}{a^2} - Q_a^3 D_{Q^3}f \right\|_2 + \|Q_a^3 D_{Q^3}f - D_{Q^3}f\|_2. \end{aligned}$$

Since the operators Q_a^3 are uniformly bounded, applying (23) we get

$$\left\| \frac{Q_a^3 Q_a^3 f - Q_a^3 f}{a^2} - Q_a^3 D_{Q^3}f \right\|_2 \leq C \left\| \frac{Q_a^3 f - f}{a^2} - D_{Q^3}f \right\|_2 \leq Ca|f|_{3,2}.$$

We now turn to the second term of (24). Applying (7) we get

$$\|Q_a^3 D_{Q^3}f - D_{Q^3}f\|_2 \leq Ca^2 |D_{Q^3}f|_{2,2}.$$

This finishes the proof of the first inequality.

The proof of the second formula is rather similar to that of [8, Theorem 1.4]. ■



Now the choice of the estimator of θ_3 is obvious:

$$\hat{\theta}_3(a) = \int_{\mathbb{R}^2} (g_{an})^2.$$

We can turn to the second strategy of estimating θ_3 . Note that

$$(25) \quad (g_{an})^2(x) = \left(\frac{1}{na^4}\right)^2 \left(\sum_{k \neq l}^n \kappa(x/a, X_k/a) \kappa(x/a, X_l/a) + \sum_{k=1}^n (\kappa(x/a, X_k/a))^2\right)$$

and

$$E(g_{an}(x))^2 = \frac{n^2 - n}{n^2} \left(\frac{T_a(f)(x)}{a^2}\right)^2 + \frac{1}{a^8} \frac{1}{n} \int_{\mathbb{R}^2} (\kappa(x/a, y/a))^2 f(y) dy.$$

Now another estimator (“no diagonals”) of θ_3 can be given by the formula

$$\hat{\theta}_3(a) = \left(\frac{1}{na^4}\right)^2 \int_{\mathbb{R}^2} \sum_{k \neq l}^n \kappa(x/a, X_k/a) \kappa(x/a, X_l/a) dx.$$

To avoid tedious calculations we propose the following simpler estimator for an even size of a sample:

$$(26) \quad \hat{\theta}_3(a) = \frac{1}{(n/2)a^8} \int_{\mathbb{R}^2} \sum_{k=1}^{n/2} \kappa(x/a, X_k/a) \kappa(x/a, X_{n-k}/a) dx.$$

This approach with minor changes is applicable to the histogram and the linear histogram. We will see it in the next section for the histogram, i.e. we will construct $\hat{\theta}_1$.

4. Choice of bandwidth for estimation of the asymptotic constant: histogram. We will explain the estimation of the asymptotic constant in the case of the histogram.

First we construct an operator Q^5 reproducing the polynomials of degree less than or equal to two by the formula

$$Q^5(f) = \sum_{|\gamma| \leq 1} a_\gamma Q^3(f(\cdot - \gamma)).$$

Applying (10) and (14) we obtain, for $|\beta| = 2$,

$$Q^3(\square^\beta)(x) - x^\beta = A_\beta,$$

where $A_{(1,1)} = 0$, $A_{(2,0)} = 1/3$, $A_{(0,2)} = 1/3$. Consequently, to find the coefficients a_γ we need to solve the system of equations, for all $|\beta| \leq 2$,

$$Q^5(\square^\beta) = \sum_{|\gamma| \leq 1} a_\gamma Q^3((\cdot - \gamma)^\beta) = \square^\beta.$$



One of the solutions is $a_{(0,0)} = 4/3$ and $a_{(\gamma_1, \gamma_2)} = -1/12$ for all $|\gamma_1| = |\gamma_2| = 1$ and the other a_γ are zero. Since Q_3 reproduces polynomials of degree less than or equal to two by (7), there is $C > 0$ such that for all $f \in W_2^3$,

$$\|Q_h^5 f - f\|_2 \leq Ch^3 |f|_{3,2}.$$

By definition,

$$Q^5(f)(x) = \int_{\mathbb{R}^2} K^5(x, y) f(y) dy,$$

where

$$K^5(x, y) = \sum_{\alpha \in \mathbb{Z}^2} F_3(y - \alpha) G_5(x - \alpha), \quad G_5(x) = \sum_{|\gamma| \leq 1} a_\gamma G_3(x - \gamma).$$

Now we consider the operator defined as follows:

$$T^1 := Q^1 \circ Q^5 - Q^5.$$

Using this operator we construct an estimator of θ_1 . We can write

$$T^1(f)(x) = \int_{\mathbb{R}^2} \kappa^1(x, y) f(y) dy$$

with

$$\kappa^1(x, y) = \sum_{\alpha \in \mathbb{Z}^2} F_3(y - \alpha) K(x - \alpha), \quad K(x) = Q^1(G_5)(x) - G_5(x).$$

Let X_1, \dots, X_n be a random sample from a distribution with density f . For simplicity let n be even. To avoid tedious calculations let

$$(27) \quad \hat{\theta}_1(a) = \frac{1}{(n/2)a^6} \int_{\mathbb{R}^2} \sum_{k=1}^{n/2} \kappa^1(x/a, X_k/a) \kappa^1(x/a, X_{n-k}/a) dx,$$

for short $\hat{\theta}_1 = \hat{\theta}_1(a)$. By definition,

$$E\hat{\theta}_1(a) = \int_{\mathbb{R}^2} \left(\frac{T_a^1 f}{a} \right)^2.$$

We have

$$E[\hat{\theta}_1 - \theta_1]^2 = E[\hat{\theta}_1 - E\hat{\theta}_1]^2 + [E\hat{\theta}_1 - \theta_1]^2.$$

The asymptotic behavior of the deterministic part follows from Theorem 4.1 and the equality

$$[E\hat{\theta}_1 - \theta_1]^2 = \left| \left\| \frac{T_a^1 f}{a} \right\|_2 - \text{Asym}(f, 1) \right|^2 + \left\| \frac{T_a^1 f}{a} \right\|_2 + \text{Asym}(f, 1) \right|^2.$$

THEOREM 4.1. *Let $f \in W_2^3$. Then*

$$\left| \left\| \frac{T_a^1 f}{a} \right\|_2 - \text{Asym}(f, 1) \right| \leq C(a|f|_{2,2} + a^2|f|_{3,2}).$$

Proof. From Theorem 8 and Lemma 11 of [9] we infer that there is $C > 0$ such that for $g \in W_2^2$,

$$\left| \left\| \frac{Q_a^1 g - g}{a} \right\|_2 - \text{Asym}(g, 1) \right| \leq C a |g|_{2,2}.$$

Put $g = Q_a^5 f$. Consequently,

$$\left| \left\| \frac{Q_a^1 Q_a^5 f - Q_a^5 f}{a} \right\|_2 - \text{Asym}(Q_a^5 f, 1) \right| \leq C h |Q_a^5 f|_{2,2}.$$

From Corollary 2.1 of [10] we obtain, for $|\beta| = 2$,

$$\|D^\beta Q_a^5 f - D^\beta f\|_2 \leq C a |f|_{3,2}.$$

Then

$$|Q_a^5 f|_{2,2} \leq C(|f|_{2,2} + a|f|_{3,2}).$$

Consequently,

$$(28) \quad \left| \left\| \frac{Q_a^1 Q_a^5 f - Q_a^5 f}{a} \right\|_2 - \text{Asym}(Q_a^5 f, 1) \right| \leq C(a|f|_{2,2} + a^2|f|_{3,2}).$$

From the triangle inequality

$$\begin{aligned} & |\text{Asym}(Q_a^5 f, 1) - \text{Asym}(f, 1)| \\ & \leq \left(\int_{\mathbb{R}^2} \left(\int_{[0,1]^2} \left| \sum_{|\beta|=1} \frac{1}{\beta!} |D^\beta f(t) - D^\beta Q_a^5 f(t)| (Q^1(\square^\beta)(x) - \square^\beta(x)) \right|^2 dx \right) dt \right)^{1/2}. \end{aligned}$$

By the above mentioned Corollary 2.1 of [10] with $|\beta| = 1$ we have

$$(29) \quad |\text{Asym}(Q_a^5 f, 1) - \text{Asym}(f, 1)| \leq C a^2 |f|_{2,2}.$$

Combining (28) and (29) we finish the proof. ■

We need the following lemma.

LEMMA 4.1. *Let $I := [0, 1] \times [1, 2]$. Let f be a bounded density such that $f \in L^1(\mathbb{R}^2) \cap L^2(\mathbb{R}^2)$. Then for fixed $\alpha \in \mathbb{Z}^2$,*

$$\lim_{h \rightarrow 0} \frac{1}{h^2} \sum_{\alpha_1 \in \mathbb{Z}^2} \int_{Ih + \alpha_1 h} f(u) du \int_{Ih + (\alpha_1 + \alpha)h} f(u) du = \int_{\mathbb{R}^2} f^2.$$

The proof is left to the reader. We can reformulate the lemma to obtain the following statement. Let f be a density such that $f \in L^2(\mathbb{R}^2)$. Let $\alpha \in \mathbb{Z}^2$ be fixed. Then for a.e. $x_1, x_2 \in [0, 1]^2$,

$$\lim_{h \rightarrow 0} \sum_{\alpha_1 \in \mathbb{Z}^2} f(hx_1 + \alpha_1 h) f(hx_2 + \alpha_1 h + \alpha h) h^2 = \int_{\mathbb{R}^2} f^2.$$

We mention this because the convergence of the Riemann sums was observed for Lebesgue-integrable functions in the papers [4] and [12].

Let us note that the support of the function F_3 is equal to I .

THEOREM 4.2. Let f be a density such that $f \in W_2^2$. If $na^6 \rightarrow \infty$ and $a \rightarrow 0$ then

$$(30) \quad \lim na^6 E(\widehat{\theta}_1 - E\widehat{\theta}_1)^2 = 2 \int_{\mathbb{R}^2} f^2 \sum_{\alpha \in \mathbb{Z}^2} b_\alpha^2,$$

where

$$b_\alpha = \int_{\mathbb{R}^2} K(x)K(x - \alpha) dx.$$

Proof. Write

$$\begin{aligned} & E(\widehat{\theta}_1 - E\widehat{\theta}_1)^2 \\ &= E\left(\frac{1}{(n/2)a^6} \int_{\mathbb{R}^2} \sum_{k=1}^{n/2} \kappa^1(x/a, X_k/a) \kappa^1(x/a, X_{n-k}/a) dx - \int_{\mathbb{R}^2} \left(\frac{T_a^1 f}{a}\right)^2\right)^2 \\ &= \frac{4}{n^2 a^4} E\left(\sum_{k=1}^{n/2} \int_{\mathbb{R}^2} \left(\frac{1}{a^4} \kappa^1(x/a, X_k/a) \kappa^1(x/a, X_{n-k}/a) - (T_a^1 f(x))^2\right) dx\right)^2 \\ &= \frac{2}{na^4} E\left(\int_{\mathbb{R}^2} \left(\frac{1}{a^4} \kappa^1(x/a, X_1/a) \kappa^1(x/a, X_2/a) - (T_a^1 f(x))^2\right) dx\right)^2 \\ &= \frac{2}{na^4} E\left(\int_{\mathbb{R}^2} \frac{1}{a^4} \kappa^1(x/a, X_1/a) \kappa^1(x/a, X_2/a) dx\right)^2 - \frac{2}{na^4} \left(\int_{\mathbb{R}^2} (T_a^1 f)^2\right)^2. \end{aligned}$$

Only the first term of the last formula is important (let us denote it by P). Using the assumption $f \in W_2^2$ we find that the second term is negligible. Using the kernel representation we get

$$\begin{aligned} P &= \frac{2}{na^4} E\left(\int_{\mathbb{R}^2} \frac{1}{a^4} \kappa^1(x/a, X_1/a) \kappa^1(x/a, X_2/a) dx\right)^2 \\ &= \frac{2}{na^4} E\left(\int_{\mathbb{R}^2} \frac{1}{a^4} \sum_{\alpha_1 \in \mathbb{Z}^2} \sum_{\alpha_2 \in \mathbb{Z}^2} F_3(X_1/a - \alpha_1) F_3(X_2/a - \alpha_2) \right. \\ &\quad \left. \times K(x/a - \alpha_1) K(x/a - \alpha_2) dx\right)^2 \\ &= \frac{2}{na^4} E\left(\frac{1}{a^2} \sum_{\alpha_1 \in \mathbb{Z}^2} \sum_{\alpha_2 \in \mathbb{Z}^2} F_3(X_1/a - \alpha_1) F_3(X_2/a - \alpha_2) b_{\alpha_1 - \alpha_2}\right)^2 \\ &= \frac{2}{na^8} \sum_{\alpha_1 \in \mathbb{Z}^2} \sum_{\alpha_2 \in \mathbb{Z}^2} \sum_{\alpha_3 \in \mathbb{Z}^2} \sum_{\alpha_4 \in \mathbb{Z}^2} E(F_3(X_1/a - \alpha_1) F_3(X_1/a - \alpha_3)) \\ &\quad \times E(F_3(X_2/a - \alpha_2) F_3(X_2/a - \alpha_4)) b_{\alpha_1 - \alpha_2} b_{\alpha_3 - \alpha_4}. \end{aligned}$$



Observe that since F_3 is the characteristic function of I , if $\alpha_1 \neq \alpha_3$ we have

$$E(F_3(X_1/a - \alpha_1)F_3(X_1/a - \alpha_3)) = 0,$$

while if $\alpha_1 = \alpha_3$,

$$\begin{aligned} E(F_3(X_1/a - \alpha_1))^2 &= \int_{\mathbb{R}^2} (F_3(u/a - \alpha_1))f(u) du \\ &= \int_{Ia+\alpha_1a} f(u) du, \end{aligned}$$

where (recall) $I = [0, 1] \times [1, 2]$. Consequently,

$$P = \frac{2}{na^8} \sum_{\alpha_1 \in \mathbb{Z}^2} \sum_{\alpha_2 \in \mathbb{Z}^2} (b_{\alpha_1 - \alpha_2})^2 \int_{Ia+\alpha_1a} f(u) du \int_{Ia+\alpha_2a} f(u) du.$$

Using Lemma 4.1 finishes the proof since $b_\alpha = 0$ for $|\alpha| > 4$. ■

Note that applying the two last theorems we deduce that to estimate the asymptotic constant the bandwidth is $a_{\text{MISE}} \sim (1/n)^{1/8}$.

5. Simulation results. We show an accuracy of the estimation of the asymptotic constant for the histogram. We take the dimension $d = 1$ and 1000 samples from the distribution of random variables

$$X = \sigma Z + 3\sigma Y,$$

where Z is standard normal $N(0, 1)$. The random variable Y is independent of Z and has binomial distribution with $p = 0.5$. We estimate

$$\theta_1 = \frac{1}{12} \int_{\mathbb{R}} (f')^2.$$

In the case of $d = 1$ the formula (27) can be written as

$$\hat{\theta}_1(a) = \frac{2}{na^3} \sum_{k=1}^{n/2} \sum_{j \in \mathbb{Z}} \sum_{|l| \leq 4} A_l I_{[1,2]}(X_n/a - j) I_{[1,2]}(X_{n-k}/a - j - l),$$

where $I_{[1,2]}$ is the characteristic function of $[1, 2]$,

$$\begin{aligned} A_l &= \int_{\mathbb{R}} K(x+l)K(x) dx, \\ K(x) &= -\frac{1}{6}(Q_1(G_3) - G_3)(x - 1) \\ &\quad + \frac{4}{3}(Q_1(G_3) - G_3)(x) - \frac{1}{6}(Q_1(G_3) - G_3)(x + 1) \end{aligned}$$

and G_3 is the B-spline, i.e.

$$G_3(x) = \frac{1}{2} \sum_{j=0}^3 (-1)^{r-j} \binom{r}{j} (j-x)_+^2.$$

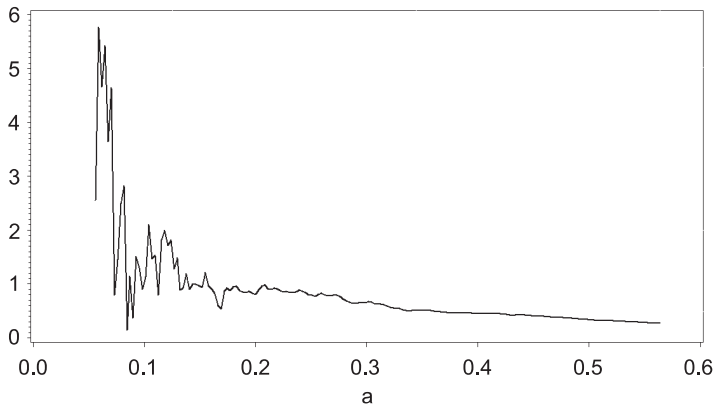
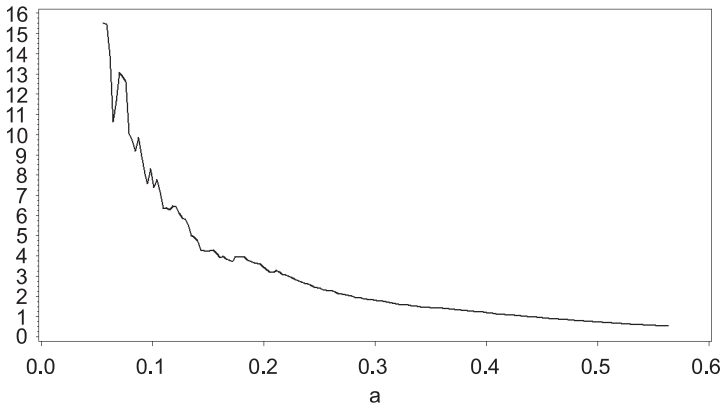
In this case, Q_1 is the orthogonal projection

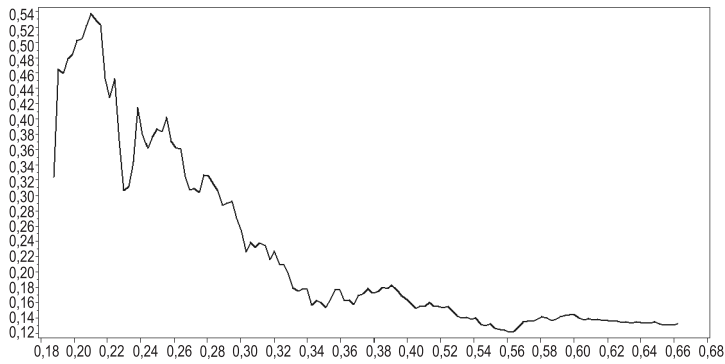
$$Q_1 f(x) = \sum_{k \in \mathbb{Z}} \int_k^{k+1} f(u) du I_{[k, k+1]}(x).$$

We have

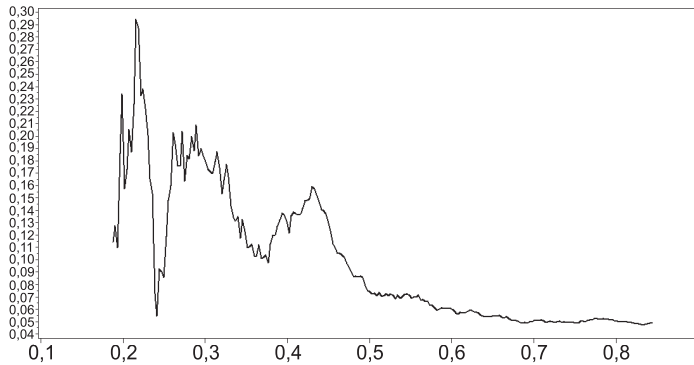
σ	0.1	0.2	0.3	0.4
θ_1	3.709	0.464	0.137	0.058

We show the four functions $\hat{\theta}_1(a)$ (SAS 9) with respect to different σ from 0.1 to 0.4. The simulations show that a for which $\hat{\theta}_1(a)$ gives a good estimation of θ_1 lies in the region where the oscillations diminish. It would be interesting to prove it.





a



a

Acknowledgments. This research is supported in part by the grant “The exploitation of white cabbage for phytoremediation and biofumigation of soils (AGROBIOKAP)” co-financed by European Union from European Regional Development Fund in a framework of the Innovative Economy Operational Programme 2007-2013.

References

- [1] A. Berlinet, *Hierarchies of higher order kernels*, Probab. Theory Related Fields 94 (1993), 489–504.
- [2] M. Beška and K. Dziedziul, *Asymptotic formula for the error in cardinal interpolation*, Numer Math. 89 (2001), 445–456.
- [3] C. de Boor, K. Höllig and S. Riemenschneider, *Box Splines*, Springer, 1993.
- [4] C. Cabrelli and U. M. Molter, *Density of the sets of generators of wavelet systems*, Constr. Approx. 26 (2007), 65–81.
- [5] Z. Ciesielski, *Asymptotic nonparametric spline density estimation in several variables*, in: Int. Ser. Numer. Math. 94, Birkhäuser, Basel, 1990, 25–53.



- [6] K. Dziedziul, *The saturation theorem for quasiprojections*, Studia Sci. Math. Hungar. 35 (1999), 99–111.
- [7] —, *Central limit theorem for square error of multivariate nonparametric box-spline density estimators*, Appl. Math. (Warsaw) 28 (2001), 437–456.
- [8] —, *Application of Mazur–Orlicz’s theorem in AMISE calculation*, ibid. 29 (2002), 33–41.
- [9] —, *Estimate of the error of approximation by asymptotic constant*, East J. Approx. 9 (2003), 385–405.
- [10] —, *Shift invariant operators and a saturation theorem*, Appl. Math. (Warsaw) 30 (2003), 267–286.
- [11] S. Efromovich and A. Samarov, *Adaptive estimation of the integral of squared regression derivatives*, Scand. J. Statist. 27 (2000), 335–351.
- [12] D. Jimenez, L. Wang and Y. Wang, *PCM quantization errors and the white noise hypothesis*, SIAM J. Math. Anal. 38 (2007), 2042–2056.
- [13] M. C. Jones, J. S. Marron and B. U. Park, *A simple root n bandwidth selector*, Ann. Statist. 19 (1991), 1919–1932.
- [14] M. C. Jones, J. S. Marron and S. J. Sheather, *Progress in data-based bandwidth selection for kernel density estimation*, Comput. Statist. 11 (1996), 337–381.
- [15] B. U. Park and J. S. Marron, *Comparison on data-driven bandwidth selectors*, J. Amer. Statist. Assoc. 85 (1990), 66–72.
- [16] S. R. Sain, K. A. Baggerly and D. W. Scott, *Cross-validation of multivariate densities*, J. Amer. Statist. Assoc. 89 (1994), 807–817.
- [17] D. W. Scott, *Multivariate Density Estimation: Theory, Practice and Visualisation*, Wiley, New York, 1991.
- [18] S. J. Sheather and M. C. Jones, *A reliable data-based bandwidth selection method for kernel density estimation*, J. Roy. Statist. Soc. B 53 (1991), 683–690.

Karol Dziedziul
Faculty of Applied Mathematics
Gdańsk University of Technology
Narutowicza 11/12
80-952 Gdańsk, Poland
E-mail: kdz@mif.pg.gda.pl

Piotr Paluszek
Institute of Mathematics
Gdańsk University
Wita Stwosza 57
80-952 Gdańsk, Poland
E-mail: peter-m-p@o2.pl

Received on 12.10.2009;
revised version on 14.9.2010

(2020)



