

MEAN-SQUARED ERROR OF KERNEL ESTIMATORS FOR FINITE VALUES OF THE SAMPLE SIZE

I. K. Glad,¹ N. L. Hjort,¹ and N. G. Ushakov²

UDC 519.2

1. Introduction

The asymptotic approach is the main tool in construction and investigation of statistical procedures (see, for example, [1]). One represents the risk of the statistical procedure under consideration as the sum of the main asymptotic term and the remainder. The main term is explicitly and relatively simply expressed in terms of the parameters of the procedure. The dependence of the remainder on the parameters and characteristics of the procedure may be quite complicated, but the remainder is smaller than the main term as the sample size tends to infinity. The remainder is then ignored and the procedure is optimized with respect to the main asymptotic term of the risk. This main term is considered as an approximation of the real risk.

Sometimes, however, the ratio of the remainder to the main asymptotic term tends to zero very slowly.

Such a situation takes place, in particular, in nonparametric kernel estimation, which is one of the most popular methods of density estimation, having a number of advantages and being quite well developed and studied (see, for example, [2, 8]). Let $X_1, \dots, X_n$ be independent and identically distributed (i.i.d.) random variables with absolutely continuous distribution function $F(x)$ and density function $f(x)$. The kernel density estimator associated with the sample $X_1, \dots, X_n$ is defined as

$$f_n(x; h) = \frac{1}{nh} \sum_{j=1}^n K\left(\frac{x - X_j}{h}\right),$$

where $K(x)$ is the kernel function with scaled version

$$K_h(x) = \frac{1}{h} K\left(\frac{x}{h}\right),$$

and $h = h_n$ is a positive number (depending on n) called the bandwidth or the smoothing parameter. The performance of kernel density estimators is usually studied via Taylor expansions and asymptotic approximation arguments, in which the bandwidth parameter tends to zero with increasing sample size. Under some standard assumptions, the mean integrated squared error (MISE) of the kernel estimator is represented as the sum of the main term and the remainder

$$\text{MISE} = \text{AMISE} + \text{REM},$$

where the AMISE (asymptotic MISE) is explicitly expressed in terms of parameters of the kernel and of the density to be estimated, and

$$\text{REM} = o(\text{AMISE}), \quad h \rightarrow 0, \quad n \rightarrow \infty.$$

Typically, the remainder is ignored and all procedures (like selection of the smoothing parameter, estimation of the real error — MISE, etc.) are based on the AMISE. But here the ratio REM/AMISE decreases (as n increases and h decreases) very slowly. This is demonstrated by Table 1, where we give values of the MISE, AMISE, REM, and ratio REM/MISE for sample sizes 20, 100, 1000, 10^4 , and 10^6 . Here both the density to be estimated and the kernel are the standard normal density. In this case, MISE, AMISE, and REM are found explicitly, namely,

$$\begin{aligned} \text{MISE}(f_n(x; h)) &= \frac{1}{2\sqrt{\pi}} \left(1 - 2\sqrt{\frac{2}{2+h^2}} + \frac{1}{\sqrt{1+h^2}} + \frac{1}{nh} - \frac{1}{n\sqrt{1+h^2}} \right), \\ \text{AMISE}(f_n(x; h)) &= \frac{3}{32\sqrt{\pi}} h^4 + \frac{1}{2\sqrt{\pi} nh}, \end{aligned}$$

and

$$R(K_h * f) = \frac{1}{2\sqrt{\pi(h^2 + 1)}}.$$

The bandwidth h is taken to be optimal. It is seen that the remainder is an essential part of the error, even for very large values of the sample size.

This paper focusses directly on the finite-sample situation. Informative upper bounds are derived for the integrated mean-squared error function and for the mean-squared error. Results are reached for the traditional case, where the kernel is “conventional,” i.e., it is a probability density function (is nonnegative and integrates to one), under various sets of assumptions on the underlying density to be estimated. Many estimates obtained in the paper are sharp.

Denote by $\varphi(t)$ and $\psi(t)$ characteristic functions, corresponding to $f(x)$ and $K(x)$, and denote by $\varphi_n(t)$ the empirical characteristic function of the sample $X_1, \dots, X_n$, that is,

$$\varphi_n(t) = \frac{1}{n} \sum_{k=1}^n e^{itX_k}.$$

TABLE 1.

n	MISE(min)	AMISE	REM	REM/MISE·100%
20	0.0158	0.0309	0.0119	74.96%
50	0.0087	0.0147	0.0050	57.63%
100	0.0054	0.0084	0.0026	47.60%
1000	0.0010	0.0013	0.0003	26.44%
10^4	$1.807E-04$	$2.100E-04$	$2.782E-05$	15.39%
10^5	$3.039E-05$	$3.329E-05$	$2.805E-06$	9.23%
10^6	$4.989E-06$	$5.276E-06$	$2.814E-07$	5.64%

2. MSE and MISE Representations

In this work, our criteria of the performance of estimators are the mean-squared error (MSE) and the mean integrated squared error (MISE), which are defined as

$$\text{MSE}(f_n(x; h)) = \mathbf{E}[f_n(x; h) - f(x)]^2$$

and

$$\text{MISE}(f_n(x; h)) = \int \text{MSE}(f_n(x; h)) dx = \mathbf{E} \int [f_n(x; h) - f(x)]^2 dx,$$

respectively. The mean-squared error and the mean integrated squared error admit well-known decompositions into squared bias (integrated squared bias) and variance (integrated variance):

$$\text{MSE}(f_n(x; h)) = \text{B}(f_n(x; h))^2 + \text{Var}(f_n(x; h))$$

and

$$\text{MISE}(f_n(x; h)) = \int \text{B}(f_n(x; h))^2 dx + \int \text{Var}(f_n(x; h)) dx, \quad (1)$$

where

$$\text{B}(f_n(x; h)) = \mathbf{E}f_n(x; h) - f(x)$$

and

$$\text{Var}(f_n(x; h)) = \mathbf{E}(f_n(x; h) - \mathbf{E}f_n(x; h))^2$$

are the bias and the variance, respectively.

The bias and the integrated squared bias are represented in the following forms:

$$\text{B}(f_n(x; h)) = \int K(z)[f(x - hz) - f(x)] dz = \frac{1}{2\pi} \int e^{-itx}[1 - \psi(ht)]\varphi(t) dt, \quad (2)$$

$$\int \text{B}(f_n(x; h))^2 dx = \frac{1}{2\pi} \int |1 - \psi(ht)|^2 |\varphi(t)|^2 dt. \quad (3)$$

Corresponding representations of the variance and integrated variance are

$$\begin{aligned} \text{Var}(f_n(x; h)) &= \frac{1}{nh} \int K(z)^2 f(x - hz) dz - \frac{1}{n} [\mathbf{E}f_n(x; h)]^2 \\ &= \frac{1}{n} \frac{1}{(2\pi)^2} \iint e^{-i(u+v)x} \psi(hu) \psi(hv) [f(u+v) - f(u)f(v)] du dv, \end{aligned} \quad (4)$$

and

$$\int \text{Var}(f_n(x; h)) dx = \frac{1}{2\pi n} \int (1 - |\varphi(t)|^2) |\psi(ht)|^2 dt. \quad (5)$$

3. Inequalities for the Mean Integrated Squared Error

For a real-valued function $g(x)$, we will use the following notations, provided that the integrals exist:

$$\mu_k(g) = \int |x|^k g(x) dx, \quad k = 0, 1, 2, \dots, \quad R(g) = \int g^2(x) dx.$$

If the density to be estimated, $f(x)$, has finite second moment, we will denote the variance of $f(x)$ by σ^2 . Denote also

$$c(h) = \min \left\{ \frac{1}{\sigma}, \frac{1}{\sqrt{\mu_2(K)} h} \right\}. \quad (6)$$

Let m be a positive integer number. Consider the following parameter of the underlying distribution:

$$P_m(\varphi) = \frac{1}{2\pi} \int t^{2m} |\varphi(t)|^m dt.$$

For even m this parameter is simply expressed in terms of the underlying density:

$$P_m(\varphi) = \int f^{(m)}(x)^2 dx = R(f^{(m)}).$$

When m is odd, $P_m(\varphi)$ is not expressed in terms of $f(x)$ explicitly but plays a similar role. In this work, we use it in the case $m = 1$. The index will be omitted, so

$$P(\varphi) = \frac{1}{2\pi} \int t^2 |\varphi(t)| dt.$$

3.1. Smooth case. In this subsection, we suppose that the density to be estimated is twice differentiable. More exactly, we suppose that the following conditions are satisfied:

- (a) the density $f(x)$ is two-times differentiable and the second derivative is square integrable,
- (b) the kernel $K(x)$ is a symmetric square-integrable probability density function with finite second moment.

Under these conditions, $\text{MISE}(f_n(x; h))$ is represented in the form

$$\text{MISE}(f_n(x; h)) = \text{AMISE}(f_n(x; h)) + o\left(h^4 + \frac{1}{nh}\right), \quad h \rightarrow 0,$$

where AMISE (asymptotic MISE) is

$$\text{AMISE}(f_n(x; h)) = \frac{1}{4} h^4 \mu_2(K)^2 R(f'') + \frac{1}{nh} R(K)$$

(see [3, 4, 8]). It turns out that a much more informative relation takes place, namely, the following theorem.

THEOREM 1. *Let conditions (a) and (b) be satisfied. Then for all n*

$$\text{MISE}(f_n(x; h)) < \frac{1}{4} h^4 \mu_2(K)^2 R(f'') + \frac{1}{nh} R(K) - \frac{c(h)}{\pi n} \left[1 - \frac{1}{3} (\sigma^2 + \mu_2(K) h^2) c(h)^2 + \frac{1}{5} \sigma^2 \mu_2(K) h^2 c(h)^4 \right], \quad (7)$$

where $c(h)$ is defined by (6).

Proof. Since $K(x)$ is symmetric, $\psi(t)$ is real; therefore, due to representations (1), (3), and (5),

$$\text{MISE}(f_n(x; h)) = \frac{1}{2\pi} \int |\varphi(t)|^2 (1 - \psi(ht))^2 dt + \frac{1}{n} \frac{1}{2\pi} \int (1 - |\varphi(t)|^2) (\psi(ht))^2 dt. \quad (8)$$

For all t we have (see [5, 6])

$$(1 - \psi(ht))^2 \leq \frac{\mu_2(K)^2 h^4 t^4}{4}.$$

Using this inequality, (8), and (several times) the Parseval–Plancherel identity, we obtain

$$\begin{aligned} \text{MISE}(f_n(x; h)) &\leq \frac{\mu_2(K)^2 h^4}{4} \frac{1}{2\pi} \int t^4 |\varphi(t)|^2 dt + \frac{1}{n} \frac{1}{2\pi} \int |\psi(ht)|^2 dt - \frac{1}{n} \frac{1}{2\pi} \int |\psi(ht)\varphi(t)|^2 dt \\ &= \frac{\mu_2(K)^2 h^4}{4} \int (f''(x))^2 dx + \frac{1}{nh} \int K(x)^2 dx - \frac{1}{n} \int [(K_h \star f)(x)]^2 dx \\ &= \frac{h^4 \mu_2(K)^2 R(f'')}{4} + \frac{1}{nh} R(K) - \frac{1}{n} R(K_h \star f). \end{aligned}$$

Let us obtain a lower bound for $(1/n)R(K_h \star f)$. The inequality

$$|\varphi(t)|^2 |\psi(ht)|^2 \geq (1 - \sigma^2 t^2)(1 - \mu_2(K)h^2 t^2)$$

holds for all t (see above); therefore,

$$\begin{aligned} R(K_h \star f) &= \frac{1}{2\pi} \int |\varphi(t)|^2 |\psi(ht)|^2 dt \geq \frac{1}{\pi} \int_0^{c(h)} (1 - \sigma^2 t^2)(1 - \mu_2(K)h^2 t^2) dt \\ &= \frac{c(h)}{\pi} \left[1 - \frac{1}{3}(\sigma^2 + \mu_2(K)h^2)c(h)^2 + \frac{1}{5}\sigma^2 \mu_2(K)h^2 c(h)^4 \right], \end{aligned}$$

and (7) follows.

It is clear that typically σ^2 is greater than $h^2 \mu_2(K)$. In this case, $c(h) = 1/\sigma$ and the last summand on the right-hand side of (7) becomes

$$\frac{1}{\pi n \sigma} \left[1 - \frac{1}{3}(\sigma^2 + \mu_2(K)h^2) \frac{1}{\sigma^2} + \frac{1}{5} \mu_2(K)h^2 \frac{1}{\sigma^2} \right],$$

which is greater than $(3\pi\sigma n)^{-1}$. Thus, we get the following simple (but quite accurate) inequality.

THEOREM 2. *Let conditions (a) and (b) be satisfied and, in addition, $\sigma^2 \geq h^2 \mu_2(K)$. Then for all n*

$$\text{MISE}(f_n(x; h)) < \frac{1}{4} h^4 \mu_2(K)^2 R(f'') + \frac{1}{nh} R(K) - \frac{1}{3\pi\sigma n}.$$

Theorem 1 immediately implies the following theorem.

THEOREM 3. *Let conditions (a) and (b) be satisfied. Then for all n*

$$\inf_{h>0} \text{MISE}(f_n(x; h)) < \frac{5}{4} [\mu_2(K)^2 R(f'') R(K)^4]^{1/5} n^{-4/5}.$$

3.2. Intermediate case. Now we consider densities having first derivative but not the second one. In this case, the order of consistency of the kernel estimator is $O(n^{-2/3})$. We obtain rigorous inequalities.

THEOREM 4. *Let $f(x)$ be differentiable and its derivative be square integrable. If $K(x)$ is a symmetric probability density function with finite first moment, then for all n ,*

$$\text{MISE}(f_n(x; h)) < h^2 \mu_1(K)^2 R(f') + \frac{1}{nh} R(K) - \frac{c(h)}{\pi n} \left[1 - \frac{1}{3}(\sigma^2 + \mu_2(K)h^2)c(h)^2 + \frac{1}{5}\sigma^2 \mu_2(K)h^2 c(h)^4 \right],$$

where $c(h)$ is defined by (6).

Proof. We have (see [6])

$$(1 - \psi(ht))^2 \leq \mu_1(K)^2 h^2 t^2.$$

Using this inequality, formulas (1), (3), (5), and (several times) the Parseval–Plancherel identity, we obtain

$$\begin{aligned} \text{MISE}(f_n(x; h)) &< h^2 \mu_1(K)^2 \frac{1}{2\pi} \int t^2 |\varphi(t)|^2 dt + \frac{1}{nh} \frac{1}{2\pi} \int |\psi(t)|^2 dt - \frac{1}{n} R(K_h \star f) \\ &< h^2 \mu_1(K)^2 R(f') + \frac{1}{nh} R(K) - \frac{1}{n} R(K_h \star f). \end{aligned}$$

The term $(1/n)R(K_h \star f)$ is estimated in the same way as in the proof of Theorem 1.

THEOREM 5. *Let the conditions of Theorem 4 be satisfied and, in addition, $\sigma^2 \geq h^2 \mu_2(K)$. Then for all n*

$$\text{MISE}(f_n(x; h)) < h^2 \mu_1(K)^2 R(f') + \frac{1}{nh} R(K) - \frac{1}{3\pi\sigma n}.$$

Theorem 5 is obtained from Theorem 4 in the same way as Theorem 2 is obtained from Theorem 1. Finally, Theorem 4 implies the following theorem.

THEOREM 6. *Let the conditions of Theorem 4 be satisfied. Then*

$$\inf_{h>0} \text{MISE}(f_n(x; h)) < \frac{3}{2^{2/3}} [\mu_1(K)^2 R(K)^2 R(f')]^{1/3} n^{-2/3}.$$

3.3. Nonsmooth case. Now we consider the nonsmooth case, that is, the densities considered in this subsection are not assumed to be differentiable or even continuous. We only suppose that they have bounded total variation, which we denote by $V(\cdot)$ and which is defined for a real-valued function $g(x)$ as

$$V(g) = \sup \sum_{i=1}^n |g(x_i) - g(x_{i-1})|,$$

where sup is taken over all n and all collections $x_0, x_1, \dots, x_n$ such that $-\infty < x_0 < x_1 < \dots < x_n < \infty$. As to the kernel, we suppose that $K(x)$ is a symmetric density function. No assumption is made about moments of the kernel.

It is convenient to use the following parameter of the kernel, which was introduced by Watson and Leadbetter [9] and which is expressed in terms of its characteristic function $\psi(t)$:

$$Q(K) = \frac{1}{2\pi} \int \left(\frac{1 - \psi(t)}{t} \right)^2 dt.$$

Note that $Q(K)$ exists (is finite) under quite mild conditions, in particular, it is sufficient if the expectation of $K(x)$ exists.

THEOREM 7. *Let $K(x)$ be a symmetric density function and $f(x)$ have bounded total variation. Then*

$$\text{MISE}(f_n(x; h)) \leq h V(f)^2 Q(K) + \frac{1}{nh} R(K). \quad (9)$$

Proof. Estimate the first summand on the right-hand side of (8) using the inequality

$$|\varphi(t)| \leq \frac{V(f)}{|t|},$$

which holds for all t (see [7]). Then

$$\frac{1}{2\pi} \int |\varphi(t)|^2 (1 - \psi(ht))^2 dt \leq V(f)^2 h \frac{1}{2\pi} \int \left(\frac{1 - \psi(t)}{t} \right)^2 dt = V(f)^2 P(K) h.$$

The second term on the right-hand side of (8) is estimated as above (see, for example, the proof of Theorem 1):

$$\frac{1}{n} \frac{1}{2\pi} \int (1 - |\varphi(t)|^2) |\psi(ht)|^2 dt \leq \frac{1}{nh} R(K).$$

Thus, we obtain (9).

Minimizing the right-hand side of (9), we obtain the following theorem.

THEOREM 8. *Let the conditions of Theorem 7 be satisfied. Then*

$$\inf_{h>0} \text{MISE}(f_n(x; h)) \leq 2V(f)[R(K)Q(K)]^{1/2}n^{-1/2}.$$

4. Inequalities for the Mean-Squared Error

Unlike for the mean integrated squared error, for which we have obtained upper bounds even better (smaller) than the main term in the asymptotic representation, the maximum of the mean-squared error does not admit upper bounds having the same order of consistency as the error if one makes no additional assumptions about the density to be estimated.

In this section, we suppose that the density to be estimated is bounded and denote

$$M(f) = \sup_x f(x).$$

THEOREM 9. *If the kernel $K(x)$ is a symmetric, square-integrable density function, then*

$$\sup_x \text{MSE}(f_n(x; h)) \leq \frac{1}{4}\mu_2(K)^2P(\varphi)^2h^4 + \frac{1}{nh}R(K)M(f). \quad (10)$$

Proof. Using representation (4) for the variance, we immediately obtain that

$$\text{Var}(f_n(x; h)) \leq \frac{1}{nh}R(K)M(f). \quad (11)$$

Let us estimate the squared bias. Making use of the inequality

$$|1 - \psi(t)| \leq \frac{1}{2}\mu_2(K)t^2,$$

which holds for all t (see above), and (2), we obtain

$$|B(f_n(x; h))| \leq \frac{1}{2\pi} \int |1 - \psi(hu)| |\varphi(u)| du = \frac{1}{2}\mu_2(K)C(\varphi)h^2.$$

Representation (10) follows from this inequality and (11).

COROLLARY. *Under the conditions of the theorem*

$$\inf_{h>0} \sup_x \text{MSE}(f_n(x; h)) \leq \frac{5}{4}[\mu_2(K)R(K)^2M(f)^2C(\varphi)]^{1/5}n^{-4/5}.$$

Remark. Naturally, the theorem has sense only when all four parameters $\mu_2(K)$, $R(K)$, $M(f)$, and $C(\varphi)$ are finite. The finiteness of the first two parameters, which concern the kernel, is the standard assumption in the “smooth case” (see subsection 3.1). The last two parameters, which concern the density to be estimated, are closely related. In particular, if the second of them is finite, then the first one is also finite. The existence of the second derivative of $f(x)$ and its square integrability do not guarantee the finiteness of $C(\varphi)$; it is necessary to make additional assumptions ($C(\varphi)$ is finite, for example, if $f(x)$ is three-times differentiable and the third derivative has finite total variation). In fact, despite condition (a) being sufficient for

$$\text{MSE}(f_n(x; h)) = O(n^{-4/5}), \quad n \rightarrow \infty,$$

this relation is not uniform over the set of densities satisfying (a). This means that there does not exist an upper bound holding for all densities of this set and having order $n^{-4/5}$.

Under milder conditions than the finiteness of $C(\varphi)$, one can obtain upper bounds having worse order of decrease than $n^{-4/5}$.

THEOREM 10. Let $K(x)$ be a symmetric square-integrable density function with finite first moment. If $f(x)$ is twice differentiable with $f''(x)$ a function of bounded variation, then

$$\sup_x \text{MSE}(f_n(x; h)) \leq \frac{9}{4\pi^2} \mu_1(K)^2 V(f)^{4/3} h^2 + \frac{1}{nh} R(K) M(f). \quad (12)$$

Proof. For the variance of $f_n(x; h)$, (11) holds. For characteristic functions $\varphi(t)$ and $\psi(t)$, we have the inequalities

$$|\varphi(t)| \leq \frac{V(f)}{|t|^3}, \quad -\infty < t < \infty,$$

(see [7]) and

$$|1 - \psi(t)| \leq \mu_1(K)ht, \quad -\infty < t < \infty,$$

(see [6]). Using representation (2), we obtain

$$\begin{aligned} |B(f_n(x; h))| &\leq \frac{1}{2\pi} \int |1 - \psi(ht)| |\varphi(t)| dt \leq \frac{1}{2\pi} \int_{|t| \leq V(f)^{1/3}} |1 - \psi(ht)| dt + \frac{1}{2\pi} \int_{|t| > V(f)^{1/3}} |1 - \psi(ht)| |\varphi(t)| dt \\ &\leq \frac{1}{2\pi} \mu_1(K)h \int_{|t| \leq V(f)^{1/3}} t dt + \frac{1}{2\pi} \mu_1(K) V(f)h \int_{|t| > V(f)^{1/3}} t^{-2} dt = \frac{3}{2\pi} \mu_1(K) V(f)^{2/3} h. \end{aligned}$$

Now (12) follows from this inequality and (11).

COROLLARY. Under the conditions of the theorem,

$$\inf_{h>0} \sup_x \text{MSE}(f_n(x; h)) \leq \frac{3^{5/3}}{(4\pi)^{2/3}} [\mu_1(K) R(K) M(f) V(f)^{2/3}]^{2/3} n^{-2/3}.$$

REFERENCES

1. N. Balakrishnan, I. A. Ibragimov, and V. B. Nevzorov, *Asymptotic Methods in Probability and Statistics with Applications*, Birkhauser, Boston (2001).
2. J. Fan and I. Gijbels, *Local Polynomial Modeling and Its Applications*, Monographs on Statistics and Applied Probability, Chapman and Hall, London (1996).
3. E. Parzen, "On the estimation of a probability density function and the mode," *Ann. Math. Statist.*, **33**, 1065–1076 (1962).
4. M. Rosenblatt, "Remarks on some nonparametric estimates of a density function," *Ann. Math. Statist.*, **27**, 832–837 (1956).
5. N. A. Sapogov, "Weak stability of J. Marcinkiewicz's theorem and some inequalities for characteristic functions," *Zapiski Nauchn. Sem. LOMI*, **85**, 193–196 (1979).
6. N. G. Ushakov, *Selected Topics in Characteristic Functions*, VSP, Utrecht (1999).
7. V. G. Ushakov and N. G. Ushakov, "Some inequalities for characteristic functions of densities with bounded variation," *Moscow Univ. Comp. Math. Cyber.*, No. 3, 45–52 (2000).
8. M. P. Wand and M. C. Jones, *Kernel Smoothing*, Chapman and Hall, London (1995).
9. G. S. Watson and M. R. Leadbetter, "On the estimation of the probability density, I," *Ann. Math. Statist.*, **34**, 480–491 (1963).

¹Department of Mathematics, University of Oslo, 1053, Blindern, NO-0316 Oslo, Norway.

²Department of Mathematical Sciences, Norwegian University of Science and Technology, 7491 Trondheim, Norway.