Some Inequalities for the Mean Integrated Squared Error of Multivariate Kernel Density Estimators

Nikolai Ushakov
Dept of Mathematical Sciences
Norwegian University of Science and Technology
N-7034 Trondheim
Norway
ushakov@math.ntnu.no

Abstract

Some upper bounds for MISE of multivariate kernel density estimators are obtained. It is shown, in particular, that under some regularity conditions, the actual error is always less than the asymptotic error.

1 Introduction

The most used measure of performance of kernel density estimators as well as the basis of the choice of the smoothing parameter, bandwidth, is the mean intergated squared error (MISE) of the estimator. Practically it is usually replaced by its asymptotic approximation. That is, MISE is represented as the sum of the main term AMISE (asymptotic MISE), having a relatively simple form, and the remainder R such that

$$\text{MISE} \sim \text{AMISE}, \quad \text{R} = o(\text{MISE}) \quad \text{as} \quad n \to \infty,$$

and then the evaluation of the actual error and the bandwidth selection are performed on the basis of AMISE.

Wand and Jones (1995) discovered that, at least for convensional (nonnegative) kernels, AMISE is always strictly greater than MISE. In addition, it turns out (see for example Glad et al., 2007) that the ratio R/MISE can tend to zero very slowly, so that the difference AMISE–MISE can be substantial even for quite large sample sizes ($10^5 - 10^6$). For moderate and small sample sizes this difference is typically so large that it seems to be unreasonable to replace MISE by AMISE. This is illustrated by Fig. 1. Here are AMISE (squares) and MISE (circles) as functions of the sample size. The bandwidth is chosen to be AMISE-optimal. Both the estimated density and the kernel are the standard normal density. Further examples can be found in Marron and Wand (1992).

This makes reasonable to try to find upper bounds for MISE lieing between MISE and AMISE. In the univariate case, a number of such inequalities was obtained in Glad et al. (2007). They can give a substantial gain. For example the upper bound for MISE, given by Theorem 1 of Glad et al. (2007) is presented in Fig. 1 (diamonds).

The problem of nonparametric density estimation arises in many areas of the reliability theory, where such bounds can be used. At the same time, in many situations distributions are multivariate. In this work, some upper bounds for MISE of multivariate kernel density estimators are derived. It is proved, in particular, that the Maron-Wand inequality (MISE<AMISE) holds also in the multivariate case.

2 Inequalities

Let $\mathbf{X}_1, ..., \mathbf{X}_n$ be independent identically distributed d-dimensional random vectors with density $f(\mathbf{x})$, $\mathbf{x} = (x_1, ..., x_d)^T \in R^d$. Throughout this article, $|\mathbf{A}|$ denotes the determinant of the square matrix $\mathbf{A}$, $\int_{R^d}$ is shorthand for $\int \cdots \int_{R^d}$ and $d\mathbf{x}$ is shorthand for $dx_1 \cdots dx_d$. The general form of the kernel estimator is (Wand and Jones, 1995)

$$f_n(\mathbf{x}; \mathbf{H}) = \frac{1}{n} \sum_{j=1}^{n} K_{\mathbf{H}}(\mathbf{x} - \mathbf{X}_j),$$

where $K_{\mathbf{H}}(\mathbf{x}) = |\mathbf{H}|^{-1/2} K(\mathbf{H}^{-1/2}\mathbf{x})$, $K(\mathbf{x})$ is a multivariate kernel, and $\mathbf{H}$ is a symmetric positive definite $d \times d$ matrix — the bandwidth matrix. In this work we will suppose that $K(\mathbf{x})$ is a symmetric probability

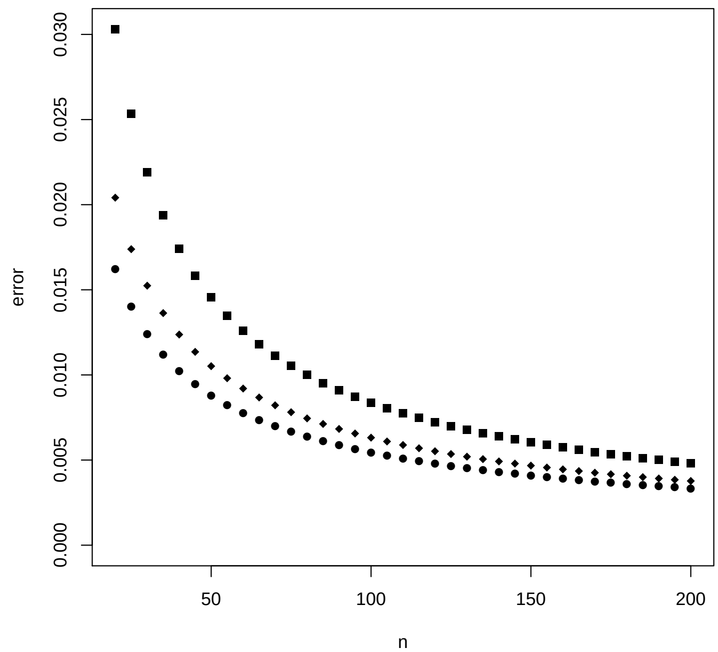

Figure 1: MISE (circles), AMISE (squares) and an upper bound for MISE (diamonds)

density function i.e. it is nonnegative, integrates to one, and $K(\mathbf{x}) = K(-\mathbf{x})$. The mean integrated squared error (MISE) of the estimator is

$$\mathrm{MISE}(f_n(\mathbf{x};\mathbf{H})) = \mathrm{E}\int_{R^d} [f_n(\mathbf{x};\mathbf{H}) - f(\mathbf{x})]^2 d\mathbf{x}.$$

Let us introduce the following conditions.

(i) All second derivatives

$$\frac{\partial^2}{\partial x_i \partial x_j} f(\mathbf{x})$$

exist and are square integrable.

(ii) The kernel $K(\mathbf{x})$ is square integrable. All second order moments of $K(\mathbf{x})$ are finite.

Denote the covariance matrices of $K(\mathbf{x})$ and $f(x)$ by $\mathbf{\Sigma}_K$ and $\mathbf{\Sigma}_f$, respectively, and entries of the matrix $\mathbf{H}^{1/2}\mathbf{\Sigma}\mathbf{H}^{1/2}$ by c_{ij}. Also, let $\mathbf{S}(\mathbf{x})$ be the Hessian matrix of the density $f(\mathbf{x})$, that is the $d \times d$ matrix having (i, j) entry equal to

$$\frac{\partial^2}{\partial x_i \partial x_j} f(\mathbf{x}).$$

Finally, denote the trace of the matrix $\mathbf{A}$ by $\mathrm{tr}\mathbf{A}$ and the integral of the squared kernel by $R(K)$:

$$R(K) = \int_{R^d} (K(\mathbf{x}))^2 d\mathbf{x}.$$

Theorem 1. *Let conditions* (i) *and* (ii) *be satisfied. Then*

$$\mathrm{MISE}(f_n(\mathbf{x};\mathbf{H})) < \frac{1}{4}\int_{R^d} \mathrm{tr}^2(\mathbf{H}^{1/2}\mathbf{\Sigma}_K\mathbf{H}^{1/2}\mathbf{S}(\mathbf{x}))d\mathbf{x} + \frac{R(K)}{n|\mathbf{H}|^{1/2}} - \frac{C(d)}{n|\mathbf{\Sigma}|^{1/2}}, \quad (1)$$

where

$$\mathbf{\Sigma} = 2(\mathbf{\Sigma}_f + \mathbf{H}^{1/2}\mathbf{\Sigma}_K\mathbf{H}^{1/2}) \tag{2}$$

and

$$C(d) = [2^{d-1}\pi^{d/2}(d+2)\Gamma(d/2+1)]^{-1}. \tag{3}$$

Suppose now that one more condition is satisfied.

(iii)

$$\int_{R^d} \mathbf{x}\mathbf{x}^T K(\mathbf{x})d\mathbf{x} = \mu_2(K)\mathbf{I}$$

where $\mu_2(K) = \int_{R^d} x_j^2 K(\mathbf{x})d\mathbf{x}$ is independent of j, and $\mathbf{I}$ is the $d \times d$ identity matrix.

Under this condition $\mathbf{H}^{1/2}\mathbf{\Sigma}_K\mathbf{H}^{1/2} = \mu_2(K)\mathbf{H}$, and from Theorem 1 we obtain the following

Theorem 2. *Let conditions* (i), (ii) *and* (iii) *be satisfied. Then*

$$\text{MISE}(f_n(\mathbf{x};\mathbf{H})) < \frac{1}{4}\mu_2(K)^2 \int_{R^d} \text{tr}^2(\mathbf{HS})d\mathbf{x} + \frac{R(K)}{n|\mathbf{H}|^{1/2}} - \frac{C(d)}{n|\mathbf{\Sigma}|^{1/2}}, \tag{4}$$

where $\mathbf{\Sigma} = 2(\mathbf{\Sigma}_f + \mu_2(K)\mathbf{H})$ *and* $C(d)$ *is given by (3).*

Corollary 1. *Let conditions* (i), (ii) *and* (iii) *be satisfied, and the bandwidth matrix depends on the sample size* n *in such a way that* $n^{-1}|\mathbf{H}|^{-1/2}$ *and all entries of the matrix* $\mathbf{H}$ *tend to zero as* $n \to \infty$. *Then*

$$\text{MISE}(f_n(\mathbf{x};\mathbf{H})) < \text{AMISE}(f_n(\mathbf{x};\mathbf{H})). \tag{5}$$

Since under conditions of the Corollary,

$$\text{AMISE}(f_n(\mathbf{x};\mathbf{H})) = \frac{1}{4}\mu_2(K)^2 \int_{R^d} \text{tr}^2(\mathbf{HS})d\mathbf{x} + \frac{R(K)}{n|\mathbf{H}|^{1/2}}$$

(Wand and Jones, 1995, p. 97), (5) immediately follows from (4).

In conclusion we consider two special, perhaps the most frequently used, smoothing parametrizations. In the first case the estimator has form

$$f_n(\mathbf{x}; h_1, ..., h_d) = \frac{1}{nh_1 \cdots h_d} \sum_{i=1}^{n} K\left(\frac{x_1 - X_{i1}}{h_1}, ..., \frac{x_d - X_{id}}{h_d}\right) \tag{6}$$

where $\mathbf{X}_i = (X_{i1}, ..., X_{id})^T$. In this case h_j can be considered as the smoothing parameter associated with the j-th coordinate direction. For this parametrization we will use only kernels with a diagonal covariance matrix. Denote

$$\mu_2^{(j)}(K) = \int_{R^d} x_j^2 K(\mathbf{x})d\mathbf{x}, \quad \sigma_j^2 = \text{Var}X_{ij}, \quad j = 1, ..., d.$$

Theorem 3. *Let conditions* (i) *and* (ii) *be satisfied, the estimator have form* (6), *and* $\mathbf{\Sigma}_K$ *be a diagonal matrix. Then*

$$\text{MISE}(f_n(\mathbf{x}; h_1, ..., h_d)) < \frac{1}{4}\int_{R^d} \left(\sum_{j=1}^{d} h_j^2 \mu_2^{(j)}(K) \frac{\partial^2 f(\mathbf{x})}{\partial x_j^2}\right)^2 d\mathbf{x} +$$

$$+ \frac{R(K)}{n}\left(\prod_{j=1}^{d} h_j\right)^{-\frac{1}{2}} - \frac{C(d)}{n\sqrt{2}}\left(\prod_{j=1}^{d}(\sigma_j^2 + \mu_2^{(j)}(K)h_j)\right)^{-\frac{1}{2}}, \tag{7}$$

where $C(d)$ *is defined by (3).*

Finally consider the following simplest parametrization. There is a single scalar smoothing parameter h and the kernel estimator is of the form

$$f_n(\mathbf{x}; h) = \frac{1}{nh} \sum_{j=1}^{n} K\left(\frac{\mathbf{x} - \mathbf{X}_j}{h}\right). \tag{9}$$

In this case we will use only kernels satisfying condition (iii)

Theorem 4. *Let conditions* (i), (ii), (iii) *be satisfied, and the estimator have form* (9)*. Then*

$$\mathrm{MISE}(f_n(\mathbf{x}; h)) < \frac{1}{4} h^4 \mu_2(K)^2 \int_{R^d} (\bigtriangledown^2 f(\mathbf{x}))^2 d\mathbf{x} + \frac{R(K)}{nh^d} - \frac{C(d)}{n\sqrt{2}} \left(\prod_{j=1}^{d} (\sigma_j^2 + \mu_2(K)h)\right)^{-\frac{1}{2}}$$

where

$$\bigtriangledown^2 f(\mathbf{x}) = \sum_{j=1}^{d} \frac{\partial^2 f(\mathbf{x})}{\partial x_j^2}.$$

References

[1] Glad, I.K., Hjort, N.L. and Ushakov, N.G. (2007). Mean-squared error of kernel estimators for finite values of the sample size. *J. Math. Sci. (N. Y.) 146*, no. 4, 5977-5983.

[2] Marron, J.S. and Wand, M.P. (1992). Exact mean integrated squared error. *Ann. Statist. 20*, 712-736.

[3] Wand, M.P. and Jones, M.C. (1995). *Kernel smoothing.* Chapman and Hall, London.