



**HAL**  
open science

## Kernel adjusted density estimation

Ramidha Srihera, Winfried Stute

► **To cite this version:**

Ramidha Srihera, Winfried Stute. Kernel adjusted density estimation. *Statistics and Probability Letters*, 2011, 81 (5), pp.571. 10.1016/j.spl.2011.01.013 . hal-00725101

**HAL Id: hal-00725101**

**<https://hal.science/hal-00725101>**

Submitted on 24 Aug 2012

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

## Accepted Manuscript

Kernel adjusted density estimation

Ramidha Srihera, Winfried Stute

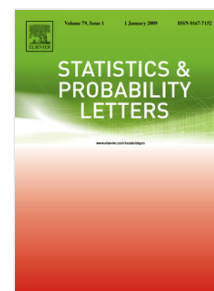
PII: S0167-7152(11)00020-4  
DOI: 10.1016/j.spl.2011.01.013  
Reference: STAPRO 5888

To appear in: *Statistics and Probability Letters*

Received date: 30 November 2010

Revised date: 18 January 2011

Accepted date: 18 January 2011



Please cite this article as: Srihera, R., Stute, W., Kernel adjusted density estimation. *Statistics and Probability Letters* (2011), doi:10.1016/j.spl.2011.01.013

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

# Kernel Adjusted Density Estimation

Ramidha Srihera, Winfried Stute

*Dept. of Mathematics and Statistics, Thammasat University, Pathum Thani, Thailand and  
Mathematical Institute, University of Giessen, Arndtstr. 2, D-35392 Giessen, Germany*

---

## Abstract

We propose and study a kernel estimator of a density in which the kernel is adapted to the data but not fixed. The smoothing procedure is followed by a location-scale transformation to reduce bias and variance. The new method naturally leads to an adaptive choice of the smoothing parameters which avoids asymptotic expansions.

*Key words:* Kernel density estimator, adaptive choice

---

*Running Title:* Kernel Adjusted Density Estimation

## 1. Introduction and Main Results

Since Rosenblatt (1956) and Parzen (1962) introduced the kernel estimator of an unknown density  $f$ , there have been numerous authors who studied various of its finite and large sample properties. To be more precise, let  $K$  be a given function on the real line, the “kernel”, and let  $h > 0$  be a given bandwidth or window size. Then, if  $X_1, \dots, X_n$  denotes an independent sample from  $f$ , the associated kernel estimator is defined as

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

To obtain a “bona fide” estimator, i.e., one which is itself a density, one has to require

$$K \geq 0 \text{ and } \int K(u)du = 1.$$

Silverman (1986) and Wand and Jones (1995) became standard reference books on kernel methodology. To cite only one of the many properties of  $f_n(x)$ , recall that for the mean square error (MSE), we have, when  $\int uK(u)du = 0$  and  $f$  is twice continuously differentiable in a neighborhood of  $x$ , that

$$\begin{aligned} \text{Bias} f_n(x) &:= \mathbb{E}f_n(x) - f(x) \\ &= \frac{1}{2}f''(x)h^2 \int u^2K(u)du + O(h^3) \end{aligned}$$

and

$$\text{Var}f_n(x) = \frac{1}{nh}f(x) \int K^2(u)du + o\left(\frac{1}{nh}\right)$$

whenever  $n \rightarrow \infty$  and  $h \rightarrow 0$  such that  $nh \rightarrow \infty$ . This implies that

$$\begin{aligned} \text{MSE}f_n(x) &= \text{Bias}^2f_n(x) + \text{Var}f_n(x) \\ &\sim \frac{1}{4}(f''(x))^2h^4 \left[ \int u^2K(u)du \right]^2 + \frac{1}{nh}f(x) \int K^2(u)du \end{aligned} \quad (1.1)$$

The optimal choice of  $h$  minimizing the last expression satisfies

$$h_{\text{opt}}^5 \sim \frac{1}{n} \frac{f(x) \int K^2(u)du}{[f''(x) \int u^2K(u)du]^2}. \quad (1.2)$$

If, rather than  $\text{MSE}f_n(x)$  at a fixed  $x$ , one considers the integrated MSE as a measure of fit, i.e.,

$$\text{MISE} = \int \text{MSE}f_n(x)dx,$$

then the optimal  $h$  satisfies, up to remainders,

$$h_{\text{opt}}^5 = \frac{1}{n} \frac{\int K^2(u)du}{\int [f''(x)]^2dx [\int u^2K(u)du]^2}. \quad (1.3)$$

It is known, see Silverman (1986), that the choice of  $K$  has little effect on MSE and MISE. Rather, the unknown  $f(x)$  and  $f''(x)$  are crucial and prevent one from a straightforward application of (1.2) or (1.3). One possibility is to choose a preliminary  $h^1$ , estimate  $f(x)$  and  $f''(x)$  and then compute an adapted version of  $h_{\text{opt}}$ . Another strategy is to determine  $h$  in a fully adaptive way by minimizing a cross-validated deviation between  $f_n$  and  $f$ . Finally, a third popular method consists in referring  $\int [f''(x)]^2dx$  in (1.3) to a standard distribution, i.e., to compute the integral for a parametric family of centered densities with scale parameter  $\sigma$ , and then to apply (1.3) with an estimated  $\sigma$ . Silverman (1986) pointed out that this method may lead to incorrect results when the reference densities are symmetric at zero but the true but unknown  $f$  is multimodal and thus typically has larger curvature relative to scale. Also, the first method is not fully satisfactory since it requires the subjective choice of a preliminary  $h^1$ . Finally, the cross-validated  $h$  is known to be asymptotically optimal but may show a poor behavior when sample size is small or moderate. See Feluch and Koronacki (1992).

It is the purpose of this paper to propose and study a fully adaptive approach which takes into account a modified version of the third method, in which the reference densities are associated with the true  $f$ . In other words, we shall consider the location scale family generated by the true  $f$ . Interestingly enough, to deal

with bias issues, it will not be necessary to incorporate estimators of  $f''$  based on preliminary choices of  $h$ . Also, we shall be able to get estimates of MSE and MISE and hence adaptive choices of the smoothing parameters.

To begin with, let  $K_0$  be a kernel from the location-scale family associated with  $f$ , i.e.,

$$K_0(u) = K_0(u, \theta, \sigma) = \sigma f(\sigma u + \theta). \quad (1.4)$$

For (1.1), with  $\theta = \mathbb{E}X$  and  $\sigma = 1$ , we then get, e.g.,

$$\text{MSE}f_n(x) \sim \frac{1}{4}(f''(x))^2 h^4 \text{Var}^2 X + \frac{1}{nh} f(x) \int f^2(u) du. \quad (1.5)$$

The interesting point about (1.5) is that the bias and variance parts contain terms which reflect both the local and global behavior of  $f$ , namely  $f''(x)$ ,  $f(x)$  and, respectively,  $\text{Var}X$  and  $\int f^2$ . Similarly for MISE.

For example, since typically  $\text{Var}X$  is small when  $f''(x)$  is large, (1.5) demonstrates that rather than choosing a fixed  $K$ , a properly chosen kernel from (1.4) may decrease the bias. The scaling factor  $\sigma$  gives us more flexibility. As will be seen later this will enable us to choose  $K_0$  so as to minimize MSE. Of course, since  $f$  in (1.4) is not available, we have to replace it by  $f_n$  from above. Hence our estimator becomes

$$\begin{aligned} \hat{f}_n(x) &\equiv \hat{f}_n(x, \theta, \sigma) = \frac{\sigma}{nh} \sum_{i=1}^n f_n \left( \sigma \frac{x - X_i}{h} + \theta \right) \\ &= \frac{\sigma}{n^2 h^2} \sum_{j=1}^n \sum_{i=1}^n K \left( \frac{\sigma x - \sigma X_i + \theta h - h X_j}{h^2} \right). \end{aligned}$$

In other words, the  $\hat{f}_n$  constitute kernel estimators with the kernels taken from the location-scale family associated with a classical kernel estimator. The choice of  $h, \sigma$  and  $\theta$  will be discussed later.

To reduce a possible bias, our final estimator will be

$$\hat{f}_n(x) = \frac{\sigma}{n(n-1)h^2} \sum_{i \neq j} K \left( \frac{\sigma x - \sigma X_i + \theta h - h X_j}{h^2} \right). \quad (1.6)$$

We first study the bias and the variance of  $\hat{f}_n(x)$  for fixed  $\theta$  and  $\sigma$ . For this, the following regularity assumptions for  $K$  and  $f$  will be required:

(A1):  $K$  is a symmetric density with compact support, i.e., satisfies

$$K(-u) = K(u) \text{ for all } u \in \mathbb{R}.$$

(A2):  $f$  is twice continuously differentiable in a neighborhood  $U$  of  $x$ .

**Theorem 1.1.** *Under (A1) and (A2), assume that  $\mathbb{E}X^2 < \infty$ . Then, if  $n \rightarrow \infty$  and  $h \rightarrow 0$  such that  $nh \rightarrow \infty$ , we have*

$$\text{Bias} \hat{f}_n(x) = \sigma^{-1} f'(x) h \int f(y)(\theta - y) dy + \frac{f''(x) h^2}{2\sigma^2} \int f(y)(\theta - y)^2 dy + O(h^3) \quad (1.7)$$

and

$$nh \text{Var} \hat{f}_n(x) = \sigma f(x) \int f^2(y) dy + o(1). \quad (1.8)$$

Both expansions hold true uniformly in  $(\theta, \sigma)$  as long as  $\theta$  ranges in a compact set and  $\sigma > 0$  is bounded away from zero.

The support condition on  $K$  is needed to exploit the local structure of the data. As a consequence, only regularity of  $f$  in a neighborhood of  $x$  is required. The assertion of Theorem 1.1 also holds for more general kernels, in which case we also need some mild regularity conditions for  $f$  outside of  $U$ . In particular, Theorem 1.1 holds if instead of (A1) and (A2) the following conditions hold:

(B1):  $K$  is a bounded symmetric density satisfying

$$\int |u|^3 K(u) du < \infty.$$

(B2):  $f$  is twice continuously differentiable in a neighborhood of  $x$ .

(B3):  $f$  is bounded and continuous on the real line.

If we set  $\theta = \mathbb{E}X$ , then the bias reduces to

$$\text{Bias} \hat{f}_n(x) = \frac{f''(x) h^2}{2\sigma^2} \text{Var} X + O(h^3).$$

Furthermore, if  $h = o(n^{-1/5})$ , the bias is negligible and the overall error is dominated by noise. In particular, we get the following result

**Theorem 1.2.** *Under the conditions of Theorem 1.1, if  $\theta = \mathbb{E}X$  and  $h = o(n^{-1/5})$ , then*

$$(nh)^{1/2} [\hat{f}_n(x) - f(x)] \rightarrow \mathcal{N}(0, \rho^2) \text{ in distribution,}$$

where

$$\rho^2(x) = \sigma f(x) \int f^2(y) dy.$$

In the following result we study the distributional convergence of  $\hat{f}_n(x)$  when unknown parameters are estimated, i.e., our estimator equals

$$\hat{f}_n(x) = \hat{f}_n(x, \hat{\theta}, \hat{\sigma}). \quad (1.9)$$

Here,  $\hat{\theta} = n^{-1} \sum_{i=1}^n X_i$  and  $\hat{\sigma}$  is a pre-specified  $\sigma$  or an estimator such that, e.g.,  $\hat{\sigma} \rightarrow \sigma$ . It is not difficult to see that as a process in  $(\theta, \sigma)$

$$(nh)^{1/2}[\hat{f}_n(x, \theta, \sigma) - f(x)]$$

is tight in the space of continuous functions, when  $\theta$  ranges in a compact set,  $\sigma$  is bounded away from zero and in addition  $K$  is continuously differentiable. From this, the following result is immediate.

**Theorem 1.3.** *Under the conditions of Theorem 1.2., if  $\mathbb{E}X^2 < \infty$  and  $\hat{\sigma} \rightarrow \sigma$  in probability, then*

$$(nh)^{1/2}[\hat{f}_n(x, \hat{\theta}, \hat{\sigma}) - f(x)] \rightarrow \mathcal{N}(0, \rho^2) \text{ in distribution,}$$

where as before

$$\rho^2(x) = \sigma f(x) \int f^2(y) dy.$$

Theorems 1.1 – 1.3 deal with  $\hat{f}_n$  at a given point  $x$ . We only mention in passing that corresponding results may also be obtained for MISE. This, however, is beyond the scope of the present paper.

## 2. Optimal Choice Of Smoothing Parameters

Theorems 1.2 and 1.3 have analogues in the classical kernel estimation literature. Choosing suboptimal  $h$  also there leads to limit distributions with expectation zero. In such a situation the scale parameter  $\sigma$  should be chosen as small as possible to make also  $\rho^2(x)$  small.

It is the purpose of this section to discuss the delicate question how to choose, in a fully adaptive way and for finite sample size, the smoothing parameter  $h$  and the scale parameter  $\sigma$  so as to minimize MSE and MISE. We only discuss MSE, the other case requiring similar arguments. Now, the optimal choice of the smoothing parameters is obtained by minimizing the sum of the leading terms of (1.7) (squared) and (1.8) at  $\theta = \mathbb{E}X$ :

$$\begin{aligned} \text{MSE} &= \text{Bias}^2 + \text{Var} \\ &= \frac{[f''(x)]^2 h^4}{4\sigma^4} \text{Var}^2 X + \frac{\sigma f(x) \int f^2(y) dy}{nh}. \end{aligned} \quad (2.1)$$

The sum depends on  $h$  and  $\sigma$  only through  $a = h/\sigma$ . It is minimal when

$$a^5 = \frac{f(x) \int f^2}{n[f''(x)\text{Var}X]^2}.$$

This may be achieved by setting  $h = n^{-1/5}$  and

$$\sigma^5 = \frac{[f''(x)\text{Var}X]^2}{f(x) \int f^2}. \quad (2.2)$$

Our last observation should be compared with what happens in the case of the classical kernel density estimator. There the optimal bandwidth satisfies, see (1.2),

$$h_{\text{opt}}^5 = \frac{f(x) \int K^2(u)du}{n[f''(x) \int u^2 K(u)du]^2} \equiv cn^{-1}$$

with the (unknown)  $c$  depending on  $f$  and  $f''$ . In our case the optimal  $h$  is specified only up to a constant so that  $h = n^{-1/5}$  is feasible. The role of the critical  $c$  is taken by the scale parameter  $\sigma$ , which though is not part of the original  $f_n$ . Hence in what follows we may set w.l.o.g.  $h = n^{-1/5}$ . Also set, as before,  $\hat{\theta} = \frac{1}{n} \sum_{i=1}^n X_i$ . Hence MSE is only a function of  $\sigma$ . To find its minimizer, we shall not, as in classical kernel density estimation, dwell on the expansion of MSE. Rather we shall employ explicit expressions of Bias and Var. As will be shown in Section 4, we have

$$\begin{aligned} \text{Bias}(\sigma) \equiv \text{Bias} &= \iint K(u)f(y)f\left(x + \frac{\theta h - hy - h^2u}{\sigma}\right) dydu - f(x) \\ &= \iint K(u)f\left(x + \frac{\theta h - hy - h^2u}{\sigma}\right) F(dy)du - f(x), \end{aligned}$$

where  $F$  denotes the distribution function pertaining to  $f$ . Furthermore, it will be shown there that the variance of  $\hat{f}_n(x)$  may be approximated by the variance of its Hájek projection  $\hat{f}_n^0(x)$ , namely

$$\sigma^2 h^{-4} n^{-1} \text{Var} \left[ \int K\left(\frac{\sigma x - \sigma X_1 + \theta h - hy}{h^2}\right) F(dy) + \int K\left(\frac{\sigma x - \sigma z + \theta h - hX_1}{h^2}\right) F(dz) \right].$$

See (4.1) below. To estimate these quantities, we introduce the empirical distribution function

$$F_n(y) = \frac{1}{n} \sum_{i=1}^n 1_{\{X_i \leq y\}}, \quad y \in \mathbb{R}.$$

Then  $\text{Bias}(\sigma)$  at  $\theta = \mathbb{E}X$  is estimated by

$$\widehat{\text{Bias}}(\sigma) = \iint K(u)f_n\left(x + \frac{\hat{\theta}h - hy - h^2u}{\sigma}\right) F_n(dy)du - f_n(x),$$



while the variance is estimated by the sample variance  $\widehat{\text{Var}}$  of the quantities

$$\int K\left(\frac{\sigma x - \sigma X_i + \hat{\theta}h - hy}{h^2}\right) F_n(dy) + \int K\left(\frac{\sigma x - \sigma z + \hat{\theta}h - hX_i}{h^2}\right) F_n(dz),$$

$i = 1, \dots, n$ . Our final choice of  $\sigma$  then is the minimizer of

$$\widehat{\text{MSE}}(\sigma) = \widehat{\text{Bias}}^2(\sigma) + \sigma^2 h^{-4} n^{-1} \widehat{\text{Var}}(\sigma). \quad (2.3)$$

For the classical kernel estimator, the variance part also presents no problems and can be estimated, for each  $h > 0$ , through sample variances. The bias, however, equals

$$\frac{1}{h} \int K\left(\frac{x-y}{h}\right) F(dy) - f(x).$$

A simple plug-in strategy as for  $\hat{f}_n$ , replacing  $F$  and  $f$  by  $F_n$  and  $f_n$ , respectively, yields  $\widehat{\text{Bias}} = 0$ . Hence the bias problem cannot be easily solved for  $f_n$ , having led researchers to (1.1). This, however, constitutes only an analytic approximation of MSE. The appearance of  $f''$  makes things even more complicated, and most of the work on kernel density estimation in the 1990's dealt with statistical estimation of the analytic approximations of MSE and MISE. See Das Gupta (2008) for a short review. As we have seen in this section, for our new estimator the expansion (2.1) is of use only to discuss the roles of  $h$  and  $\sigma$ . The relevant question of how to choose  $\sigma$  in a fully adaptive way does not take into account any analytic approximations but relies on purely statistical arguments yielding (2.3).

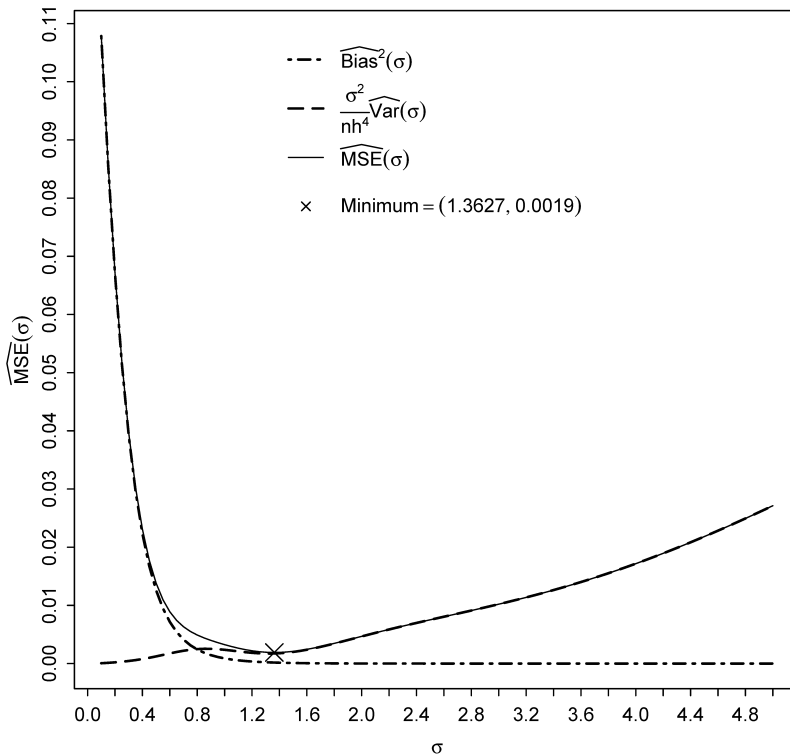
It is also of interest to compare the two expressions for MSE, i.e., (1.1) and (2.1), after an affine transformation of  $X$  to a new scale, say  $X \rightarrow \alpha X + \beta$ ,  $\alpha > 0$ . The resulting variable then has density  $\alpha^{-1} f(\frac{x-\beta}{\alpha})$ . In (1.1), on the new scale,  $(f''(x))^2$  now needs to be replaced by  $\alpha^{-6} (f''(\frac{x-\beta}{\alpha}))^2$ , inflating the contribution of the bias part when  $\alpha < 1$ . Compared with that, in (2.1), since  $\text{Var}(\alpha x + \beta) = \alpha^2 \text{Var} X$ , both parts of MSE depend on  $\alpha$  in the same way, since they are both functions of  $\alpha^{-2}$ .

### 3. A Simulation Study

In this section we report on a small simulation study which is designed to show that the minimizer of  $\widehat{\text{MSE}}(\sigma)$  in (2.3) yields a reliable choice of  $\sigma$ . In each case the data came from a standard normal density  $f$ . Also the smoothing kernel  $K$  was a standard normal density. Finally, in each case we set  $x = 0$ . As argued in the previous section, for the smoothing parameter  $h$  we may set  $h = n^{-1/5}$ . In a simulation study, when we know the true  $f$ , we may compute the true bias and variance and hence MSE. As a consequence the true  $h_{\text{opt}}$  minimizing  $\text{MSE} f_n(0)$  is available. Alternatively, we may compute our new estimator  $\hat{f}_n$  for  $h = n^{-1/5}$  and

with  $\sigma$  varying along the positive real line. Figure 1 below depicts, for a selected sample, the ingredients of (2.3). We see that  $\widehat{\text{Bias}}^2(\sigma)$  decreases as  $\sigma$  increases, while the variance part increases with  $\sigma$ . As a result, the graph of  $\widehat{\text{MSE}}\hat{f}_n(0)$  as a function of  $\sigma$  is a slightly disturbed convex function.

Figure 1: The ingredients of  $\widehat{\text{MSE}}$  as a function of  $\sigma$   
 $(n, x) = (20, 0)$



In the table to follow we compare, for different sample sizes, the optimal MSE of  $f_n(0)$ , which is unknown in a real world situation, with  $\widehat{\text{MSE}}\hat{f}_n(0)$  obtained from minimizing (2.3). Actually, the tables contain summary statistics of  $\widehat{\text{Bias}}(\sigma)$ ,  $\widehat{\text{Var}}(\sigma)$  and  $\widehat{\text{MSE}}(\sigma)$  obtained from  $M = 1000$  replications of samples of size  $n$ . It becomes clear that our adaptive choice of  $\hat{f}_n(0)$  leads to estimators whose quality is close to that of the kernel estimator with optimal but unknown bandwidth.

Table 1: Comparison of  $f_n(0)$  and  $\hat{f}_n(0)$ 

	Bias $f_n(0)$	Bias $\hat{f}_n(0)$	Var $f_n(0)$	Var $\hat{f}_n(0)$	MSE $f_n(0)$	MSE $\hat{f}_n(0)$
$n = 20$	-0.0426	-0.0512	0.0042	0.0068	0.0060	0.0094
$n = 50$	-0.0304	-0.0316	0.0024	0.0029	0.0033	0.0039
$n = 100$	-0.0244	-0.0248	0.0015	0.0018	0.0021	0.0025
$n = 200$	-0.0188	-0.0187	0.0010	0.0013	0.0014	0.0016

#### 4. Lemmas and Proofs

*Proof of (1.7).* From (1.6), we have

$$\mathbb{E}\hat{f}_n(x) = \sigma h^{-2} \iint K\left(\frac{\sigma x - \sigma z + \theta h - hy}{h^2}\right) f(y)f(z)dydz.$$

Putting

$$u = \frac{\sigma x - \sigma z + \theta h - hy}{h^2},$$

we obtain, upon noticing that  $\int uK(u)du = 0$ ,

$$\begin{aligned} \mathbb{E}\hat{f}_n(x) &= \iint K(u)f(y)f\left(x + \frac{\theta h - hy - h^2u}{\sigma}\right) dydu \\ &= \iint K(u)f(y)\left[f(x) + f'(x)\frac{\theta h - hy - h^2u}{\sigma} + \frac{1}{2}f''(x)\left(\frac{\theta h - hy - h^2u}{\sigma}\right)^2\right] dydu + O(h^3) \\ &= f(x) + \sigma^{-1}f'(x)h \iint K(u)f(y)(\theta - y)dydu \\ &\quad + \frac{f''(x)h^2}{2\sigma^2} \iint K(u)f(y)(\theta - y)^2dydu + O(h^3) \\ &= f(x) + \sigma^{-1}f'(x)h \int f(y)(\theta - y)dy + \frac{f''(x)h^2}{2\sigma^2} \int f(y)(\theta - y)^2dy + O(h^3), \end{aligned}$$

whence the assertion.  $\square$

To study the variance and the distributional behavior of  $\hat{f}_n(x)$ , we introduce its Hájek projection  $\hat{f}_n^0(x)$  defined as

$$\begin{aligned} \hat{f}_n^0(x) &= \frac{\sigma}{nh^2} \sum_{i=1}^n \left[ \int K\left(\frac{\sigma x - \sigma X_i + \theta h - hy}{h^2}\right) f(y)dy + \int K\left(\frac{\sigma x - \sigma z + \theta h - hX_i}{h^2}\right) f(z)dz \right. \\ &\quad \left. - \int \int K\left(\frac{\sigma x - \sigma z + \theta h - hy}{h^2}\right) f(z)f(y)dzdy \right]. \end{aligned}$$

Note that  $\hat{f}_n^0(x)$  is a sum of i.i.d. random variables with

$$\mathbb{E}\hat{f}_n^0(x) = \mathbb{E}\hat{f}_n(x).$$

Its variance equals

$$\begin{aligned} \text{Var} \hat{f}_n^0(x) &= \sigma^2 h^{-4} n^{-1} \mathbb{E} \left\{ \int K \left( \frac{\sigma x - \sigma X_1 + \theta h - hy}{h^2} \right) f(y) dy + \int K \left( \frac{\sigma x - \sigma z + \theta h - hX_1}{h^2} \right) f(z) dz \right. \\ &\quad \left. - 2 \int \int K \left( \frac{\sigma x - \sigma z + \theta h - hy}{h^2} \right) f(z) f(y) dz dy \right\}^2. \end{aligned} \quad (4.1)$$

In the following Lemma we provide, after a proper scaling, the limit of  $\text{Var} \hat{f}_n^0(x)$ .

**Lemma 4.1.** *As  $h \rightarrow 0$  such that  $nh \rightarrow \infty$ , we have*

$$nh \text{Var} \hat{f}_n^0(x) = \sigma f(x) \int f^2(y) dy + o(1).$$

Lemma 4.1 will be an easy consequence of Lemmas 4.3 - 4.5. Together with Lemma 4.2., this will provide the proof of (1.8) and Theorem 1.2. First we show that  $\hat{f}_n(x)$  and  $\hat{f}_n^0(x)$  are, after a proper scaling, asymptotically equivalent.

**Lemma 4.2.** *Under the assumptions of Theorem 1.1, we have*

$$nh \mathbb{E}[\hat{f}_n(x) - \hat{f}_n^0(x)]^2 \rightarrow 0 \text{ as } n \rightarrow \infty, h \rightarrow 0 \text{ such that } nh \rightarrow \infty.$$

*Proof.* By definition of  $\hat{f}_n$  and  $\hat{f}_n^0$ , we have

$$\begin{aligned} \hat{f}_n(x) - \hat{f}_n^0(x) &= \frac{\sigma}{n(n-1)h^2} \sum_{i \neq j} \left[ K \left( \frac{\sigma x - \sigma X_i + \theta h - hX_j}{h^2} \right) \right. \\ &\quad - \int K \left( \frac{\sigma x - \sigma X_i + \theta h - hy}{h^2} \right) f(y) dy \\ &\quad - \int K \left( \frac{\sigma x - \sigma z + \theta h - hX_j}{h^2} \right) f(z) dz \\ &\quad \left. + \iint K \left( \frac{\sigma x - \sigma z + \theta h - hy}{h^2} \right) f(z) f(y) dz dy \right] \\ &\equiv \frac{\sigma}{n(n-1)h^2} \sum_{i \neq j} H_h(X_i, X_j). \end{aligned}$$

It is readily seen that the last sum is a degenerate  $U$ -statistic of degree two, i.e.,

$$\mathbb{E}[H_h(X_i, X_j) H_h(X_k, X_j) | X_j] = 0$$

and

$$\mathbb{E}[H_h(X_i, X_j) H_h(X_i, X_k) | X_i] = 0$$

for  $i \neq j \neq k$ . Conclude that

$$\mathbb{E}[\hat{f}_n(x) - \hat{f}_n^0(x)]^2 = \frac{\sigma^2}{n(n-1)h^4} \mathbb{E}H_h^2(X_1, X_2).$$

It follows from arguments similar to those used for (1.7) that each of the four terms in  $H_h$  admits a second moment of the order  $O(h^2)$ . Hence

$$nh\mathbb{E}[\hat{f}_n(x) - \hat{f}_n^0(x)]^2 = O((nh)^{-1}) = o(1).$$

□

**Lemma 4.3.** *We have, as  $h \rightarrow 0$ ,*

$$\begin{aligned} h^{-3} \int \left[ \int K \left( \frac{\sigma x - \sigma z + \theta h - hy}{h^2} \right) f(y) dy \right]^2 f(z) dz \\ \rightarrow \sigma^{-1} f(x) \int f^2(y) dy. \end{aligned}$$

*Proof.* The integral equals

$$\iiint K \left( \frac{\sigma x - \sigma z + \theta h - hy_1}{h^2} \right) K \left( \frac{\sigma x - \sigma z + \theta h - hy_2}{h^2} \right) f(y_1) f(y_2) f(z) dy_1 dy_2 dz.$$

After putting

$$u = \frac{\sigma x - \sigma z + \theta h - hy_1}{h^2} \quad z = x + \frac{\theta h - hy_1 - uh^2}{\sigma}$$

it becomes

$$\frac{h^2}{\sigma} \iiint K(u) K \left( u + \frac{y_1 - y_2}{h} \right) f(y_1) f(y_2) f \left( x + \frac{\theta h - hy_1 - uh^2}{\sigma} \right) dy_1 dy_2 du.$$

Another substitution

$$v = \frac{y_1 - y_2}{h} \quad y_2 = y_1 - hv$$

leads to

$$\frac{h^3}{\sigma} \iiint K(u) K(u+v) f(y_1) f(y_1 - hv) f \left( x + \frac{\theta h - hy_1 - uh^2}{\sigma} \right) dy_1 dv du.$$

As  $h \rightarrow 0$ , the last integral tends to

$$\iiint K(u) K(u+v) f^2(y_1) f(x) dy_1 dv du = f(x) \int f^2(y) dy.$$

This completes the proof. □

**Lemma 4.4.** *As  $h \rightarrow 0$ , we have*

$$\begin{aligned} h^{-4} \int \left[ \int K \left( \frac{\sigma x - \sigma z + \theta h - hy}{h^2} \right) f(z) dz \right]^2 f(y) dy \\ \rightarrow \sigma^{-2} f^2(x). \end{aligned}$$

*Proof.* The integral equals

$$\iiint K \left( \frac{\sigma x - \sigma z_1 + \theta h - hy}{h^2} \right) K \left( \frac{\sigma x - \sigma z_2 + \theta h - hy}{h^2} \right) f(z_1) f(z_2) f(y) dz_1 dz_2 dy.$$

Putting

$$v = \frac{\sigma x - \sigma z_1 + \theta h - hy}{h^2} \quad \tilde{v} = \frac{\sigma x - \sigma z_2 + \theta h - hy}{h^2}$$

the integral becomes

$$\begin{aligned} \sigma^{-2} h^4 \iiint K(v) K(\tilde{v}) f \left( x + \frac{\theta h - hy - vh^2}{\sigma} \right) f \left( x + \frac{\theta h - hy - \tilde{v}h^2}{\sigma} \right) f(y) dv d\tilde{v} dy \\ \sim \sigma^{-2} h^4 \iiint K(v) K(\tilde{v}) f^2(x) f(y) dv d\tilde{v} dy = \sigma^{-2} h^4 f^2(x), \end{aligned}$$

whence the assertion.  $\square$

**Lemma 4.5.** *As  $h \rightarrow 0$ , we have*

$$\begin{aligned} h^{-4} \iiint K \left( \frac{\sigma x - \sigma u + \theta h - hy}{h^2} \right) K \left( \frac{\sigma x - \sigma z + \theta h - hu}{h^2} \right) f(y) f(z) f(u) du dy dz \\ \rightarrow \sigma^{-2} f^2(x). \end{aligned}$$

*Proof.* Substituting

$$\tilde{w} = \frac{\sigma x - \sigma z + \theta h - hu}{h^2},$$

the above integral becomes

$$\sigma^{-1} h^2 \iiint K \left( \frac{\sigma x - \sigma u + \theta h - hy}{h^2} \right) K(\tilde{w}) f(y) f(u) f \left( x + \frac{\theta h - hu - h^2 \tilde{w}}{\sigma} \right) du dy d\tilde{w}.$$

The substitution

$$\tilde{v} = \frac{\sigma x - \sigma u + \theta h - hy}{h^2}$$

leads to

$$\begin{aligned} \sigma^{-2} h^4 \iiint K(\tilde{v}) K(\tilde{w}) f(y) f \left( x + \frac{\theta h - hy - h^2 \tilde{v}}{\sigma} \right) f \left( x + \frac{\theta h - hu(\tilde{v}) - h^2 \tilde{w}}{\sigma} \right) dy d\tilde{v} d\tilde{w} \\ \sim \sigma^{-2} h^4 f^2(x). \end{aligned}$$

*Proof of Lemma 4.1.* It follows from the proof of (1.7) that the third integral in (4.1) is of the order  $h^2$ . Conclude from Lemmas 4.3 - 4.5 that □

$$\begin{aligned} nh\text{Var}\hat{f}_n^0(x) &\sim \sigma^2 h^{-3} \mathbb{E} \left\{ \int K \left( \frac{\sigma x - \sigma X_1 + \theta h - hy}{h^2} \right) f(y) dy \right\}^2 \\ &\rightarrow \sigma f(x) \int f^2(y) dy. \end{aligned}$$

*Proof of (1.8).* By Hájek's (1968) lemma, □

$$nh \left[ \text{Var}\hat{f}_n(x) - \text{Var}\hat{f}_n^0(x) \right] = nh \mathbb{E} \left[ \hat{f}_n(x) - \hat{f}_n^0(x) \right]^2.$$

The conclusion therefore follows from Lemma 4.1 and Lemma 4.2 . □

Theorem 1.2 follows from the fact that under  $h = o(n^{-1/5})$  the bias is negligible compared with the variance part. Secondly  $\hat{f}_n(x) - \hat{f}_n^0(x) = o((nh)^{-1/2})$  by Lemma 4.2. Hence it suffices to show asymptotic normality for  $\hat{f}_n^0(x)$ . But this constitutes a sum of i.i.d. random variables, to which the CLT applies. The limit variance follows from Lemma 4.1. This concludes the proof of Theorem 1.2.

## 5. Conclusions

Classical kernel estimation faces the problem of estimating and minimizing the MSE. This is why one often considers an analytic approximation which itself needs to be approximated. In the present paper we propose and study a new kernel estimator, which avoids these problems and allows for a direct statistical analysis of MSE without using analytic expansions. We study the large sample properties of our estimator, discuss the adaptive choice of smoothing parameters and show in a small simulation study, that the methodology is reliable already for small sample sizes.

## Acknowledgement

The authors are grateful to Dr. Gerrit Eichner for his assistance with the simulations. Thanks also to two referees for their helpful comments on an earlier version.

**References**

- Das Gupta, A. (2008). *Asymptotic theory of statistics and probability*. Springer Verlag.
- Feluch, W. & Koronacki, J. (1992). A note on modified cross-validation in density estimation. *Comp. Stat. and Data Analysis* **13**, 143-151.
- Hájek, J. (1968). Asymptotic normality of simple linear rank statistics under alternatives. *Ann. Math. Statist.* **39**, 325-343.
- Parzen, E. (1962). On the estimation of a probability density function and the mode. *Ann. Math. Statist.* **33**, 1065-1076.
- Rosenblatt, M. (1956). Remarks on some nonparametric estimates of a density function. *Ann. Math. Statist.* **27**, 832-837.
- Silverman, B. W. (1986). *Density Estimation for Statistics and Data Analysis*. Chapman & Hall, London.
- Wand, M. P. & Jones, M. C. (1995). *Kernel Smoothing*. Chapman & Hall, London.

Ramidha Srihera  
Thammasat University, Rangsit  
Center  
Department of Mathematics and  
Statistics  
Faculty of Science and Technol-  
ogy  
Pathum Thani 12121  
Thailand  
E-mail:  
ramidha@mathstat.sci.tu.ac.th

Winfried Stute  
Mathematical Institute  
University of Giessen  
Arndtstr. 2  
D-35392 Giessen  
Germany  
E-mail:  
Winfried.Stute@math.uni-  
giessen.de