South African Statistical Journal
Vol. 54, No. 1, 15–23
https://doi.org/10.37920/sasj.2020.54.1.2

EFFICIENCY BEHAVIOUR OF KERNEL-SMOOTHED KERNEL DISTRIBUTION FUNCTION ESTIMATORS

Paul Janssen[1]
Hasselt University, Center for Statistics, Agoralaan, 3590 Diepenbeek, Belgium
e-mail: *paul.janssen@uhasselt.be*

Jan W. H. Swanepoel
North-West University, Potchefstroom, South Africa
e-mail: *jan.swanepoel@nwu.ac.za*

Noël Veraverbeke
Hasselt University, Center for Statistics, Agoralaan, 3590 Diepenbeek, Belgium
North-West University, Potchefstroom, South Africa
e-mail: *noel.veraverbeke@uhasselt.be*

The asymptotic mean integrated squared error (AMISE) and the kernel efficiency (KE) of kernel distribution function estimators are well studied. In this note we define new non-parametric distribution function estimators by kernel-smoothing an initial kernel distribution function estimator. We show that, under certain conditions, the AMISE and the KE can be improved. A concrete example and a Monte Carlo simulation are worked out for illustration.

Key words: Asymptotic mean integrated squared error, Kernel distribution function estimator, Kernel efficiency.

1. Introduction

Given a random sample $X_1, \ldots, X_n$ from an unknown absolutely continuous distribution function (df) $F(\cdot)$, the kernel distribution function estimator is given by

$$\widehat{F}_h(x; K) = \frac{1}{n} \sum_{i=1}^{n} K\left(\frac{x - X_i}{h}\right), \tag{1}$$

with $K(x) = \int_{-\infty}^{x} k(t)dt$, where k is a kernel density, and bandwidth h.

In this note we apply kernel-smoothing with kernel k_1 (and corresponding $df\, K_1$) and bandwidth b to $\widehat{F}_h(\cdot; K_0)$ with initial kernel density k_0 (and corresponding $df\, K_0$) and bandwidth h, i.e., we consider

$$\widehat{F}_{h,b}(x; K_0, K_1) = \frac{1}{b} \int_{-\infty}^{+\infty} k_1\left(\frac{x - y}{b}\right) \widehat{F}_h(y; K_0)dy. \tag{2}$$

[1]Corresponding author.
MSC2010 subject classifications. 62G07, 62G20.

Note that we can rewrite $\widehat{F}_{h,b}(\cdot;K_0,K_1)$ in two ways

$$\widehat{F}_{h,b}(x;K_0,K_1) = \frac{1}{n}\sum_{i=1}^{n}\int_{-\infty}^{\infty} k_1(u)K_0\left(\frac{x-X_i}{h} - \frac{b}{h}u\right)du \tag{3}$$

and

$$\widehat{F}_{h,b}(x;K_0,K_1) = \frac{1}{n}\sum_{i=1}^{n}\int_{-\infty}^{\infty} k_0(u)K_1\left(\frac{x-X_i}{b} - \frac{h}{b}u\right)du. \tag{4}$$

From expression (3) it is clear that if $b/h \cong 0$ (i.e., $b/h \to 0$) then $\widehat{F}_{h,b}(x;K_0,K_1) \cong \widehat{F}_h(x;K_0)$. Similarly, if $h/b \cong 0$ (i.e., $h/b \to 0$), it follows from (4) that $\widehat{F}_{h,b}(x;K_0,K_1) \cong \widehat{F}_b(x;K_1)$. Therefore, the interesting case is $b/h = c$ for some constant $c > 0$. We have

$$\widehat{F}_{h,ch}(x;K_0,K_1) = \frac{1}{nch}\sum_{i=1}^{n}\int_{-\infty}^{\infty} k_1\left(\frac{x-y}{ch}\right)K_0\left(\frac{y-X_i}{h}\right)dy. \tag{5}$$

A simple calculation shows that

$$\widehat{F}_{h,ch}(x;K_0,K_1) = \frac{1}{n}\sum_{i=1}^{n} K_c\left(\frac{x-X_i}{h}\right), \tag{6}$$

which is a standard kernel distribution function estimator with K_c given by the convolution of K_0 and $k_{1,c}(y) = c^{-1}k_1(y/c)$, i.e.,

$$K_c(t) = \int_{-\infty}^{\infty} k_{1,c}(y)K_0(t-y)dy = (K_0 * k_{1,c})(t).$$

The main objective of the paper is to show that, based on a comparison of the asymptotic mean integrated squared error (AMISE) and the kernel efficiency (KE) of $\widehat{F}_{h,ch}(\cdot;K_0,K_1)$ and $\widehat{F}_h(\cdot;K_0)$, it can be beneficial to kernel-smooth a kernel distribution function estimator.

The present study parallels results in Janssen et al. (2019) on kernel-smoothed kernel density estimators.

2. Preliminaries

Throughout this paper we assume the following standard conditions on the distributions F, K_0 and K_1, and on the bandwidth h.

(F) F has two continuous derivatives f and f' and $0 < D(f) := \int_{-\infty}^{\infty}(f'(x))^2 dF(x) < \infty$.

(K) K_0 and K_1 have densities k_0 and k_1 on $(-\infty,+\infty)$ which are symmetric around 0, with $\mu_2(k_i) = \int_{-\infty}^{\infty} z^2 k_i(z)dz < \infty$, $i = 0,1$.

(h) $h = h_n$ is a sequence of positive numbers, such that $h_n \to 0$ and $nh_n \to \infty$ as $n \to \infty$.

Extensions to kernels with compact support $[-L, L]$ are possible. The results remain true but the proof of Lemma 1 is different.

Under (F), (K) and (h) a typical accuracy measure of $\widehat{F}_h(\cdot;K)$ is the AMISE, which is the leading term (n large, h small) of the mean integrated squared error (Azzalini, 1981; Altman and Leger, 1995; Swanepoel and Van Graan, 2005). It is given by

$$AMISE(\widehat{F}_h(\cdot;K)) = \frac{1}{6n} - 2\frac{h}{n}I(K)R(f) + \frac{1}{4}h^4\mu_2^2(k)D(f), \tag{7}$$

with

$$I(K) = \int_{-\infty}^{\infty} zk(z)K(z)dz = \frac{1}{2}\int_{-\infty}^{\infty} K(z)(1-K(z))dz$$

and, for every square integrable g,

$$R(g) = \int_{-\infty}^{\infty} g^2(x)dx.$$

Note that the version of MISE we are using is $E\{\int_{-\infty}^{\infty}[\widehat{F}_h(x;K) - F(x)]^2 dF(x)\}$ (see e.g. Swanepoel, 1988).

Minimising $AMISE(\widehat{F}_h(\cdot;K))$ with respect to h gives

$$h_0 = \left\{\frac{2I(K)R(f)}{\mu_2^2(k)D(f)}\right\}^{1/3} n^{-1/3}$$

as optimal bandwidth, for which

$$AMISE(\widehat{F}_{h_0}(\cdot;K)) = \frac{1}{6n} - \frac{3}{4^{1/3}}\frac{(I(K))^{4/3}}{(\mu_2(k))^{2/3}}\frac{((R(f))^{4/3}}{(D(f))^{1/3}}n^{-4/3}. \tag{8}$$

The factor in (8) that depends on the kernel K is denoted by

$$C(K) = \frac{(I(K))^{4/3}}{(\mu_2(k))^{2/3}}$$

and needs to be maximized. Applying the Cauchy–Schwarz inequality to $I(K) = \int_{-\infty}^{\infty} zk(z)(K(z) - \frac{1}{2})dz$ gives that $I(K) \leq (\mu_2(k)/12)^{1/2}$ and hence $C(K) \leq (\sqrt{12})^{-4/3}$.

This upper bound corresponds to the value of $C(K_u)$, where K_u is the df of a uniform distribution on $[-L, L]$, for any $L > 0$. See also Swanepoel (1988) and Jones (1990). Hence K_u maximizes $C(K)$ and hence minimizes $AMISE(\widehat{F}_{h_0}(\cdot;K))$ in (8) and is therefore the optimal kernel.

In terms of $C(K)$, kernel efficiency for the kernel distribution K is therefore defined as

$$KE(K) = \left\{\frac{C(K)}{C(K_u)}\right\}^{3/4}.$$

Some popular kernels with their KE values are: uniform (KE = 1.000), Epanechnikov (KE = 0.995), triangular (KE = 0.989) and normal (KE = 0.977).

In Section 3 we show that, under some conditions, $AMISE(\widehat{F}_{h,ch}(\cdot;K_0,K_1))$ is smaller than $AMISE(\widehat{F}_h(\cdot;K_0)$, i.e., kernel-smoothing the kernel distribution function is beneficial. Kernel efficiencies, $KE(K_0)$ and $KE(K_c)$, are compared in Section 4. A concrete example and a Monte Carlo simulation illustrate the theory. We conclude this section with a key lemma on the behaviour of $I(K_c)$. The technical proof is in the Appendix.

Lemma 1. *Assume* (F)*,* (K)*,* (h)*,* k_0 *has a bounded, continuous derivative almost everywhere (a.e.) with respect to Lebesgue measure which is ultimately monotone in both tails and* $R(k_0) < \infty$*. Then, for* $c \to 0$*,*

$$I(K_c) = I(K_0) + \frac{1}{2}c^2\mu_2(k_1)R(k_0) + o(c^2).$$

3. Minimized AMISE comparison

For ease of notation we write

$$\Theta_{h,0} := AMISE(\widehat{F}_h(\cdot; K_0)),$$
$$\Theta_{h,0,1}(c) := AMISE(\widehat{F}_{h,ch}(\cdot; K_0, K_1)).$$

Theorem 1. *Assume the conditions of Lemma 1. Then there exists a constant* $c_0 = c_0(k_0)$ *such that*

$$\inf_{h>0} \Theta_{h,0,1}(c) < \inf_{h>0} \Theta_{h,0} \text{ for all } 0 < c \le c_0 \tag{9}$$

if and only if

$$I(K_0) < R(k_0)\mu_2(k_0). \tag{10}$$

Proof. From (8) we have with $k_c = K_c'$,

$$\inf_{h>0} \Theta_{h,0,1}(c) - \inf_{h>0} \Theta_{h,0} = \frac{3}{4^{1/3}} \left\{ \frac{I(K_0)^{4/3}}{\mu_2(k_0)^{2/3}} - \frac{I(K_c)^{4/3}}{\mu_2(k_c)^{2/3}} \right\} \frac{(R(f))^{4/3}}{(D(f))^{1/3}} n^{-4/3},$$

and this is strictly negative if and only if

$$I(K_0)^2\mu_2(k_c) < I(K_c)^2\mu_2(k_0). \tag{11}$$

Since $\mu_2(k_c)$ is the variance of a convolution, we have that $\mu_2(k_c) = \mu_2(k_0) + c^2\mu_2(k_1)$. Using this and Lemma 1 we obtain after some easy algebra that (11) holds for all c sufficiently small if and only if (10) holds. ∎

4. Kernel efficiency

To compare the kernel efficiencies $KE(K_c)$ and $KE(K_0)$ for given kernel densities k_0 and k_1, we use the following relation, obtained from the definitions of $KE(K)$ and $C(K)$ in Section 2:

$$\frac{KE(K_0)}{KE(K_c)} = \frac{I(K_0)}{\mu_2(k_0)^{1/2}} \Bigg/ \frac{I(K_c)}{\mu_2(k_c)^{1/2}}.$$

From (11) in the proof of Theorem 1, we obtain the following result, showing that kernel-smoothing the kernel distribution function can improve the kernel efficiency.

Corollary 1. *Assume the conditions of Lemma 1. Then there exists a constant* $c_0 = c_0(k_0)$ *such that*

$$KE(K_c) > KE(K_0) \text{ for all } 0 < c \le c_0$$

if and only if

$$I(K_0) < \mu_2(k_0)R(k_0).$$

Remark 1. For K_0 the uniform on $[-L, L]$, we have $I(K_0) = L/6$, $R(k_0) = 1/(2L)$, $\mu_2(k_0) = L^2/3$ and hence we have equality in (10). So the kernel efficiency cannot be improved. This is, of course, in line with the optimality of the uniform kernel described in Section 1.

Example 1 (Convolution of Laplace and uniform kernels). Let $k_0(t) = \exp(-|t|)/2$, $-\infty < t < \infty$, and $k_{1,c}(t) = I(-c \le t \le c)/(2c)$, with $I(\cdot)$ the indicator function. Calculations show that $K_0(t) = \exp(t)/2$, for $-\infty < t \le 0$, and $K_0(t) = 1 - \exp(-t)/2$, for $0 \le t < \infty$. Also, $\mu_2(k_0) = 2$, $R(k_0) = 1/4$ and $I(K_0) = 3/8$, which imply that $\mu_2(k_0)R(k_0) > I(K_0)$, so that condition (10) of Theorem 1 is satisfied.
A direct calculation yields, for $c > 0$,

$$K_c(t) = (K_0 * k_{1,c})(t) = \begin{cases} \frac{1}{4c}\{e^{(t+c)} - e^{(t-c)}\}, & \text{for } t \le -c \\ \frac{1}{2c}(t+c) + \frac{1}{4c}\{e^{-(t+c)} - e^{(t-c)}\}, & \text{for } -c \le t \le c \\ 1 - \frac{1}{4c}\{e^{(c-t)} - e^{-(c+t)}\}, & \text{for } t \ge c. \end{cases}$$

Since $\mu_2(k_c) = \mu_2(k_0) + c^2\mu_2(k_1)$, we have

$$\mu_2(k_c) = 2 + \frac{c^2}{3}, \tag{12}$$

and after some tedious calculations we find that

$$I(K_c) = \left(\frac{c}{6} + \frac{1}{2c} - \frac{5}{16c^2}\right) + e^{-2c}\left(\frac{1}{8c} + \frac{5}{16c^2}\right). \tag{13}$$

From (12) and (13) it easily follows that $C(K_c)$ is strictly increasing as $c \to \infty$ with limit $\lim_{c\to\infty} C(K_c) = (1/12)^{2/3} \cong 0.19078$. For example, if $c \cong 16.9$ then $C(K_{16.9}) \cong 0.1907$. From the strict monotonicity of $C(K_c)$ we conclude from the proof of Theorem 1 that

$$\inf_{h>0} \Theta_{h,0,1}(c) < \inf_{h>0} \Theta_{h,0} \text{ for all } c > 0.$$

While the conclusion of Theorem 1 is only valid for c small, it turns out to be valid for all $c > 0$ in this particular example. This follows from the exact calculations above and intuitively from the fact that $K_c(t)$ approaches the uniform distribution on $[-c, c]$ when c becomes large.

Suppose we define the AMISE-efficiency of K_c with respect to K_0 (see (8)) by

$$AME(K_c) := \frac{1/(6n) - \inf_{h>0}\Theta_{h,0,1}(c)}{1/(6n) - \inf_{h>0}\Theta_{h,0}} = \frac{C(K_c)}{C(K_0)}.$$

Since $C(K_0) = (3/8)^{4/3}/2^{2/3}$ we have that $AME(K_{16.9}) \cong 1.12$. This illustrates the beneficial effect of kernel smoothing applied to the original distribution function estimator.

An interesting property of the family of kernels $\{K_c = K_0 * k_{1,c}, c > 0\}$ is that one can choose at least one of its members such that $KE(K_c) \cong 1 = KE(K_u)$. We indeed have $C(K_u) = (1/12)^{2/3}$, and hence if, for example $c = 16.9$, we obtain

$$KE(K_c) = \left\{\frac{C(K_c)}{C(K_u)}\right\}^{3/4} = 0.999663,$$

which improves

$$KE(K_0) = \left\{ \frac{C(K_0)}{C(K_u)} \right\}^{3/4} = 0.918558$$

in a substantial way.

Remark 2. For standard normal kernels $k_0(t) = k_1(t) = \phi(t)$,

$$k_c(t) = k_0 * k_{1,c}(t) = \frac{1}{\sqrt{1+c^2}} \phi\left(\frac{t}{\sqrt{1+c^2}}\right).$$

Hence $C(K_c) = C(K_0)$, $C(\cdot)$ is scale invariant, and therefore $\inf_{h>0} \Theta_{h,0,1}(c) = \inf_{h>0} \Theta_{h,0}$, i.e., we have equality in (9). Also, $AME(K_c) = 1$ for all $c \geq 0$. The reason for this is that (10) in Theorem 1 (and Corollary 1) is not satisfied, in fact we have

$$I(K_0) = \mu_2(k_0)R(k_0) = \frac{1}{2\sqrt{\pi}}.$$

This provides a nice counterexample for Theorem 1.

Remark 3. In order to make $\widehat{F}_{h,ch}(x; K_0, K_1)$ suitable for practical applications we propose, in view of the expression derived for the optimal bandwidth h_0 and the expression for $AMISE$ in (8), the following choices of the tuning parameter c and the smoothing parameter h for specified kernels K_0 and K_1:

$$c_m := \arg\max C(K_c),$$

and a simple normal-reference plug-in bandwidth selector is

$$\widehat{h}_n(c_m) := \widehat{\sigma}(108\pi)^{1/6}\{I(K_{c_m})/\mu_2^2(k_{c_m})\}^{1/3} n^{-1/3},$$

where (see Silverman, 1986, p.47)

$$\widehat{\sigma} := \min\{S, IQR/1.349\},$$

with S the sample standard deviation and IQR the interquartile range. The factor 108π is obtained by noting that $R(\phi) = (4\pi)^{-1/2}$ and $D(\phi) = \int_{-\infty}^{\infty} (\phi'(x))^2 \phi(x) dx = \sqrt{3}/(18\pi)$.

To demonstrate that the proposed plug-in bandwidth selector is effective, we perform, for $k_0(\cdot)$ the Laplace kernel and $k_{1,c}(\cdot)$ the uniform kernel on $[-c, c]$ (as in Example 1), a small Monte Carlo simulation (based on 10000 simulation runs). Table 1 displays the Monte Carlo estimate of the expected value of the ratio

$$R(n, c) = \frac{\Theta_{\widehat{h}_n(c),0,1}(c)}{\Theta_{\widehat{h}_n(0),0}},$$

with $\widehat{h}_n(c) = \widehat{\sigma}(108\pi)^{1/6}\{I(K_c)/\mu_2^2(k_c)\}^{1/3} n^{-1/3}$, for $c \geq 0$. The standard errors of the Monte Carlo estimates were found to be negligibly small and are therefore not reported in Table 1.

The grid of c-values is 0.25, 0.5, 5, 10, 20. As sample size we take $n = 20, 30, 50, 100$ and as target distributions (TD) we take

(i) $\Phi(t)$, $t \in \mathbb{R}$ (standard normal),

Table 1. Monte Carlo estimates of the expected values of $R(n, c)$ for the target distributions (c = 0.25, 0.5, 5, 10, 20 and n = 20, 30, 50, 100).

n/c	0.25	0.5	5	10	20
(i) Normal distribution					
20	0.9985711	0.9946520	0.9312509	0.9253148	0.9241872
30	0.9988013	0.9955135	0.9423249	0.9373449	0.9363990
50	0.9990366	0.9963942	0.9536471	0.9496448	0.9488845
100	0.9992814	0.9973104	0.9654247	0.9624393	0.9618723
(ii) Contaminated normal distribution					
20	0.9996937	0.9988535	0.9852620	0.9839895	0.9837478
30	0.9997861	0.9991995	0.9897089	0.9888203	0.9886515
50	0.9998678	0.9995053	0.9936402	0.9930911	0.9929867
100	0.9999107	0.9996657	0.9957030	0.9953320	0.9952615
(iii) Skewed unimodal distribution, case 1					
20	0.9985482	0.9945664	0.9301498	0.9241186	0.9229730
30	0.9987823	0.9954425	0.9414131	0.9363545	0.9353936
50	0.9990239	0.9963468	0.9530370	0.9489820	0.9482118
100	0.9992734	0.9972804	0.9650396	0.9620210	0.9614476
(iv) Skewed unimodal distribution, case 2					
20	0.9984089	0.9940450	0.9234480	0.9168381	0.9155825
30	0.9986813	0.9950644	0.9365521	0.9310737	0.9300331
50	0.9989554	0.9960902	0.9497387	0.9453989	0.9445745
100	0.9992299	0.9971176	0.9629459	0.9597464	0.9591387

Table 2. $R(f)$ and $D(f)$ for the target distributions.

TD	$R(f)$	$D(f)$
(i)	0.28209	0.03063
(ii)	0.25215	0.02407
(iii)	0.25090	0.01863
(iv)	0.19294	0.00564

(ii) $\frac{9}{10}\Phi(t) + \frac{1}{10}\left(\frac{1}{3}\Phi\left(\frac{t}{3}\right)\right)$, $t \in \mathbb{R}$ (contaminated normal),

(iii) $\frac{1}{2}\Phi(t) + \frac{1}{2}\Phi(t-1)$, $t \in \mathbb{R}$ (skewed unimodel distribution, case 1), and

(iv) $\frac{1}{2}\Phi(t) + \frac{1}{2}\Phi(t-2)$, $t \in \mathbb{R}$ (skewed unimodel distribution, case 2).

To calculate $R(n, c)$ we need for the distributions (i)–(iv): $R(f)$ and $D(f)$. They are given in Table 2.

The results in this table clearly illustrate the beneficial effect, especially for moderate sample sizes, on the AMISE when using $\widehat{F}_{h,ch}(x, K_0, K_1)$ in the estimation of normal mixture distribution functions.

Acknowledgements. We thank the associate editor and the reviewers for their valuable comments (leading e.g. to Remark 2) and good suggestions (leading e.g. to the Monte Carlo study) to improve the original manuscript.

Appendix

Proof of Lemma 1. Applying a Taylor expansion and the symmetry of k_1 around zero, we obain that

$$K_c(t) = K_0(t) + \left(\frac{c^2}{2}\right)\int_{-\infty}^{+\infty} z^2 k_0'(\eta_c(z,t))k_1(z)dz, \tag{A.1}$$

for some intermediate point $\eta_c(z, t)$ between $t - cz$ and t. For arbitrary large $T > 0$, define for any distribution function K with density function k,

$$I_T(K) = \left(\frac{1}{2}\right)\int_{-T}^{+T} K(t)(1 - K(t))dt \tag{A.2}$$

and

$$R_T(k) = \int_{-T}^{+T} k^2(t)dt. \tag{A.3}$$

Thus, using (A.1) and (A.2), we may write

$$I_T(K_c) = I_T(K_0) - \left(\frac{c^2}{2}\right)\int_{-T}^{+T} K_0(t)B_c(t)dt + \left(\frac{c^2}{4}\right)\int_{-T}^{+T} B_c(t)(dt) - \left(\frac{c^4}{8}\right)\int_{-T}^{+T} B_c^2(t)dt, \tag{A.4}$$

where

$$B_c(t) = \int_{-\infty}^{+\infty} z^2 k_0'(\eta_c(z,t))k_1(z)dz. \tag{A.5}$$

Note that since k_0' is a.e. continuous, $k_0'(\eta_c(z, t)) \to k_0'(t)$ a.e. as $c \to 0$. Also, the boundedness of k_0' implies that $|z^2 k_0'(\eta_c(z, t))k_1(z)| \leq Cz^2 k_1(z)$, for some finite constant $C > 0$, with $\int_{-\infty}^{+\infty} z^2 k_1(z)dz = \mu_2(k_1) < \infty$.

By applying Lebesgue's dominated convergence theorem we therefore have from (A.5) that, for $c \to 0$,

$$B_c(t) \to k_0'(t)\mu_2(k_1) \text{ a.e.} \tag{A.6}$$

Define, for $T > 0$,

$$J_c(T) := (I_T(K_c) - I_T(K_0))/c^2. \tag{A.7}$$

From (A.3), (A.4), (A.6) and (A.7) it follows, by using the symmetry of k_0 and K_0, and a similar argument as in the derivation of (A.6), that

$$J(T) := \lim_{c\to 0} J_c(T) = -\frac{1}{2}\mu_2(k_1)(2K_0(T) - 1)k_0(T) + \frac{1}{2}\mu_2(k_1)R_T(k_0). \tag{A.8}$$

For all T sufficiently large (A.8) yields that

$$J'(T) = -\frac{1}{2}\mu_2(k_1)(2K_0(T) - 1)k_0'(T) \geq 0,$$

since $k_0'(T) \leq 0$. This implies that $J(T)$ is nondecreasing for all large T (ultimately monotone). Thus, $J(T) \to \ell$ as $T \to \infty$, for some constant $\ell > 0$. From (A.8) it is clear that

$$J(T) \leq \frac{1}{2}\mu_2(k_1)R_T(k_0) \leq \frac{1}{2}\mu_2(k_1)R(k_0). \tag{A.9}$$

Finally, we conclude from (A.8) and (A.9) that

$$\ell = \sup_{T>0} J(T) = J(\infty) = \frac{1}{2}\mu_2(k_1)R(k_0). \tag{A.10}$$

The proof of the lemma now follows from (A.7) and (A.10). ■

References

Altman, N. and Leger, C. (1995). Bandwidth selection for kernel distribution function estimation. *Journal of Statistical Planning and Inference*, **46**, 195–214.

Azzalini, A. (1981). A note on the estimation of a distribution function and quantiles by a kernel method. *Biometrika*, **68**, 326–328.

Janssen, P., Swanepoel, J. W. H., and Veraverbeke, N. (2020). A note on the behaviour of a kernel-smoothed kernel density estimator. *Statistics & Probability Letters*, **158**. doi:10.1016/j.spl.2019.108663.

Jones, M. C. (1990). The performance of kernel density functions in kernel distribution function estimation. *Statistics & Probability Letters*, **9**, 129–132.

Silverman, B. W. (1986). *Density Estimation for Statistics and Data Analysis*, volume 26. CRC Press, Boca Raton.

Swanepoel, J. W. H. (1988). Mean intergrated squared error properties and optimal kernels when estimating a diatribution function. *Communications in Statistics – Theory and Methods*, **17**, 3785–3799.

Swanepoel, J. W. H. and Van Graan, F. C. (2005). A new kernel distribution function estimator based on a non-parametric transformation of the data. *Scandinavian Journal of Statistics*, **32**, 551–562.

Manuscript received 2019-05-27, revised 2019-12-04, accepted 2020-01-22.