Ann. Inst. Statist. Math.
Vol. 44, No. 3, 579–591 (1992)

INTEGRATED SQUARED ERROR OF KERNEL-TYPE ESTIMATOR OF DISTRIBUTION FUNCTION

SHINGO SHIRAHATA[1] AND IN-SUN CHU[2]

[1] *Department of Statistics, School of General Education, Osaka University, Toyonaka, Osaka 560, Japan*
[2] *Department of Mathematical Science, Faculty of Engineering Science, Osaka University, Toyonaka, Osaka 560, Japan*

(Received November 5, 1990; revised August 28, 1991)

Abstract. Let $X_1, \ldots, X_n$ be a random sample drawn from distribution function $F(x)$ with density function $f(x)$ and suppose we want to estimate $F(x)$. It is already shown that kernel estimator of $F(x)$ is better than usual empirical distribution function in the sense of mean integrated squared error. In this paper we derive integrated squared error of kernel estimator and compare the error with that of the empirical distribution function. It is shown that the superiority of kernel estimators is not necessarily true in the sense of integrated squared error.

Key words and phrases: Nonparametric distribution function estimator, kernel function, mean integrated squared error, integrated squared error.

1. Introduction

Let $X_1, \ldots, X_n$ be a random sample drawn from unknown continuous distribution function (d.f.) $F(x)$ with density function $f(x)$. The kernel density estimate with appropriate kernel function $k(t)$ and smoothing parameter h $(= h_n)$

$$\hat{f}_n(x) = \frac{1}{nh} \sum_{i=1}^{n} k\left(\frac{x - X_i}{h}\right)$$

is a popular nonparametric estimate of $f(x)$ which is introduced by Rosenblatt (1956) and Parzen (1962).

In the case of d.f. $F(x)$, the most natural estimator is the empirical d.f.

$$F_n(x) = \frac{1}{n} \sum_{i=1}^{n} I(x - X_i),$$

where $I(x) = 1$ (0) for $x \geq 0$ (< 0). But as is stated in Falk (1983), $F_n(x)$ does not take into account the smoothness of $F(x)$, i.e., the existence of a density $f(x)$.

Therefore many authors proposed kernel-type estimator $\hat{F}_n(x)$ of the form

$$\hat{F}_n(x) = \int_{-\infty}^{x} \hat{f}_n(y)dy = \frac{1}{n}\sum_{i=1}^{n} K_h(x - X_i),$$

where

$$K(x) = \int_{-\infty}^{x} k(y)dy \quad \text{and} \quad K_h(x) \equiv K\left(\frac{x}{h}\right).$$

Throughout this paper, we assume that the kernel function k and K satisfy

(A) $k(t)$ is a probability density such that it is bounded, symmetric around zero ($k(t) = k(-t)$) and has finite support. Thus, $k(t)$ satisfies

$$\int_{-\infty}^{\infty} k(t)dt = 1, \quad \int_{-\infty}^{\infty} tk(t)dt = 0 \quad \text{and} \quad 0 < \int_{-\infty}^{\infty} t^2 k(t)dt \equiv \kappa_2 < \infty.$$

The smoothing parameter h which tends to 0 as n tends to ∞ and the function k (or K) are to be chosen by the user. As pointed out by many authors, the choice of the kernel k is not very crucial but the choice of the smoothing parameter is a serious problem that has been addressed in the literature extensively. We also assume that the underlying d.f. $F(x)$ satisfies

(B) $F(x)$ is twice continuously differentiable with bounded $f'(x)$ and $\text{Var}(f^2(X)) < \infty$.

Several properties of $\hat{F}_n(x)$ have been investigated. Nadaraya (1964), Winter (1973) and Yamato (1973) proved the uniform convergence of $\hat{F}_n(x)$ to $F(x)$ with probability one, Watson and Leadbetter (1964) proved the asymptotic normality of $\hat{F}_n(x)$, and Winter (1979) showed that $\hat{F}_n(x)$ has the Chung-Smirnov property. Moreover, Reiss (1981) proved that the relative deficiency of the empirical d.f. $F_n(x)$ with respect to an appropriately chosen $\hat{F}_n(x)$ quickly tends to infinity as the sample size increases. Azzalini (1981) derived also an asymptotic expression (as n tends to ∞) for the mean squared error of $\hat{F}_n(x)$ and determined the asymptotically optimal smoothing parameter. See also Mack (1984) and Hill (1985).

Swanepoel (1988), under the conditions (A) and (B), derived

$$\begin{aligned}\int_{-\infty}^{\infty} & E\{\hat{F}_n(x) - F(x)\}^2 dF(x) \\ &= \frac{1}{6n} - 2hn^{-1}C\int_{-\infty}^{\infty} f^2(x)dx + \frac{1}{4}h^4\kappa_2^2\int_{-\infty}^{\infty}(f'(x))^2 f(x)dx \\ &\quad + o(hn^{-1} + h^4),\end{aligned}$$

where C is defined as

$$C \equiv \int_{-\infty}^{\infty} tk(t)K(t)dt.$$

He also proved the uniform kernel is optimal and derived optimal smoothing parameter in the sense of minimizing the mean integrated squared error (MISE) and asserted that the kernel-type estimator $\hat{F}_n(x)$ is asymptotically more efficient than the empirical d.f. $F_n(x)$ because the constant C is positive for many kernel functions. Jones (1990) supports him. Falk (1983) also stands on the same assertion for some other classes of kernels in the sense of asymptotic deficiency, and Mammitzsch (1984) investigated the problem of maximizing the constant C. Objections by practitioners against the use of kernel-type estimates mainly refer to the fact that differences are mathematically found on second order level (deficiency level), which might be hard to be realized in practice for not too large sample sizes.

The previous works referred in the above adopted MISE criterion or some other quadratic criterion, and the smoothing parameter and the kernel function are selected in view of minimizing MISE. In the present paper, we consider more basic criterion; integrated squared error (ISE) of kernel-type estimator $\hat{F}_n(x)$

$$\mathrm{ISE}(\hat{F}_n) \equiv \int_{-\infty}^{\infty} \{\hat{F}_n(x) - F(x)\}^2 dF(x)$$

by comparing with that of the empirical d.f. $F_n(x)$

$$\mathrm{ISE}(F_n) \equiv \int_{-\infty}^{\infty} \{F_n(x) - F(x)\}^2 dF(x).$$

It is proved that there is no essential difference between $\hat{F}_n(x)$ and $F_n(x)$ in the sense of ISE when estimating a continuous distribution function. Though the empirical d.f. is not continuous, many practitioners are doubtful of the worth of smoothing for empirical d.f. Our results assert that we can get no special gains from smoothing when we restrict that an estimator should be a distribution function. We further proved that we can find better estimator $\hat{F}_n(x)$ than $F_n(x)$ in the sense of ISE among some other class of kernel $k(t)$ in which $k(t)$ is not a density function (and hence $\hat{F}_n(x)$ is not a distribution function). Thus, we can conclude as follows. If a practitioner wants to estimate $F(x)$ by a distribution function, then it is enough to adopt empirical d.f. $F_n(x)$. If one believes that the estimator must minimize squared error and does not adhere to the distribution function property, then one should adopt a kernel-type estimator.

Several lemmas and the main results will be stated in Section 2. The proofs will be given in Section 3.

2. ISE of kernel-type estimator $\hat{F}_n(x)$

In this section, a comparison of the two estimators $\hat{F}_n(x)$ and $F_n(x)$ is made by comparing ISE. At first, we consider the ISE of the empirical d.f. $F_n(x)$

$$\begin{aligned}(2.1)\quad & n\,\mathrm{ISE}(F_n) \\ &= \frac{1}{n}\sum_{i=1}^{n}\int_{-\infty}^{\infty}\{I(x-X_i) - F(x)\}^2 dF(x) \\ &\quad + \frac{1}{n}\sum_{i\neq j}\int_{-\infty}^{\infty}\{I(x-X_i) - F(x)\}\{I(x-X_j) - F(x)\} dF(x) \\ &\equiv J_1 + J_2.\end{aligned}$$

Similarly, the ISE of kernel-type estimator $\hat{F}_n(x)$ is given by

$$\begin{aligned}
&n\,\mathrm{ISE}(\hat{F}_n)\\
&= n\int_{-\infty}^{\infty}[\hat{F}_n(x)-E\{\hat{F}_n(x)\}]^2dF(x)\\
&\quad + 2n\int_{-\infty}^{\infty}[\hat{F}_n(x)-E\{\hat{F}_n(x)\}][E\{\hat{F}_n(x)\}-F(x)]dF(x)\\
&\quad + n\int_{-\infty}^{\infty}[E\{\hat{F}_n(x)\}-F(x)]^2dF(x)\\
&= \frac{1}{n}\sum_{i=1}^{n}\int_{-\infty}^{\infty}\{K_h(x-X_i)-L(x)\}^2dF(x)\\
&\quad + \frac{1}{n}\sum_{i\neq j}\int_{-\infty}^{\infty}\{K_h(x-X_i)-L(x)\}\{K_h(x-X_j)-L(x)\}dF(x)\\
&\quad + 2\sum_{i=1}^{n}\int_{-\infty}^{\infty}\{K_h(x-X_i)-L(x)\}\{L(x)-F(x)\}dF(x)\\
&\quad + n\int_{-\infty}^{\infty}\{L(x)-F(x)\}^2dF(x)\\
&\equiv I_1+I_2+I_3+I_4,
\end{aligned}$$

where $L(x)=E\{K_h(x-X)\}$.

The terms I_1, I_2, I_3 and I_4 will be studied in detail, and then compared with the case of empirical d.f. Hall (1984) derived a similar expression of ISE of kernel density estimators. However, there does not exist an estimator which is the standard in density estimate and hence he did not perform any comparisons.

The proofs of following Lemmas 2.1–2.3 will be given in Section 3.

LEMMA 2.1. *Under the conditions* (A) *and* (B) *of Section* 1, I_1 *follows asymptotically a normal distribution with the parameters*

$$E(I_1)=\frac{1}{6}-2Ch\int_{-\infty}^{\infty}f^2(x)dx+o(h)$$

and

$$\begin{aligned}
\mathrm{Var}(I_1)&=\frac{1}{180n}+Cn^{-1}h\int_{-\infty}^{\infty}\left[-\frac{2}{3}+4F(x)\{1-F(x)\}\right]f(x)dF(x)\\
&\quad+o(n^{-1}h).
\end{aligned}$$

From the proof of Lemma 2.1 in Section 3, the connection with I_1 relative to empirical d.f. $F_n(x)$ is given by

$$\begin{aligned}
I_1&=J_1-2Cn^{-1}h\sum f(X_i)+O_p(h^2)\\
&=J_1-2Ch\int_{-\infty}^{\infty}f^2(x)dx+O_p(n^{-1/2}h+h^2).
\end{aligned}\tag{2.2}$$

Now let us consider the term I_2. The kernel of the term I_2, since I_2 is a degenerated U-statistic which has variable kernel function and mean 0, is given by

$$\int_{-\infty}^{\infty} \{K_h(u-x) - L(u)\}\{K_h(u-y) - L(u)\}dF(u) - G(x,y) + G(x,y)$$
$$\equiv J(x,y) + G(x,y),$$

where $J_2 \equiv (1/n)\sum_{i\neq j} G(X_i, X_j)$ which was given by (2.1).

Note that $J(x,y)$ depends on sample size n, and is also a degenerated kernel. Then

$$I_2 = \frac{1}{n}\sum_{i\neq j} J(X_i, X_j) + \frac{1}{n}\sum_{i\neq j} G(X_i, X_j) \equiv I_{21} + I_{22},$$

and by the relation (2.1),

$$I_{22} = J_2. \tag{2.3}$$

We wish to evaluate I_{21}. The problem is to evaluate

$$\int_{-\infty}^{\infty} K_h(u-x)K_h(u-y)dF(u).$$

Recall that $k(t) = 0$ for $|t| > \delta$, for some $\delta > 0$, i.e., K has finite support on $[-\delta, \delta]$. We then have

LEMMA 2.2. *Under the conditions* (A) *and* (B) *of Section* 1,

$$I_{21} = O_p(h),$$

and hence by (2.3),

$$I_2 = J_2 + O_p(h). \tag{2.4}$$

Remark 1. The uniform kernel when estimating a d.f., by Swanepoel (1988) and Jones (1990), is the optimal density on a finite interval, although it is not the best kernel density for density estimation. If the kernel is uniform on the unit interval, i.e.,

$$K(x) = \begin{cases} 0, & \text{for } x \le -\frac{1}{2} \\ x + \frac{1}{2}, & \text{for } -\frac{1}{2} < x < \frac{1}{2} \\ 1, & \text{for } x \ge \frac{1}{2}, \end{cases}$$

then $\int_{-1/2}^{1/2} K(x)dx = 1/2$. Hence, by the proof (iv) of Lemma 2.2 below, we have

$$\int_{-\infty}^{\infty} J^2(x,y)dF(x)F(y) = 2\int_{x>y} J^2(x,y)dF(x)F(y) = O(h^3)$$

and

$$I_2 = J_2 + O_p(h^{3/2}).$$

LEMMA 2.3. *Under the conditions* (A) *and* (B) *of Section* 1,

$$I_3 \sim N\left(0, \frac{\kappa_2^2 n h^4}{4} \mathrm{Var}\{f^2(X)\}\right)$$
$$\sim \frac{1}{2}\kappa_2 n^{1/2} h^2 \{\mathrm{Var}(f^2(X))\}^{1/2} N(0,1) = O_p(n^{1/2}h^2).$$

For the evaluation of I_4, by Swanepoel (1988, Lemmas 4.3–4.6), we have

LEMMA 2.4. *Under the conditions* (A) *and* (B) *of Section* 1,

$$I_4 = O(nh^4).$$

Consequently, from (2.2), (2.4), Lemmas 2.3 and 2.4, we have

THEOREM 2.1. *Under* (A) *and* (B), *the* ISE *of* $\hat{F}_n(x)$ *is given by*

$$n\,\mathrm{ISE}(\hat{F}_n) = n\,\mathrm{ISE}(F_n) - 2Ch\int_{-\infty}^{\infty} f^2(x)dx + n^{1/2}h^2\mathcal{N}$$
$$+ O_p(h) + o_p(h) + O(nh^4),$$

where $\mathcal{N}$ *is a random variable following* $N(0,1)$ *and the expectation of* $O_p(h)$ *term is zero.*

Remark 2. Suppose $h = O(n^{-1/2+a})$ for some $a > 0$. If $1/6 < a < 1/4$, then the preceding Theorem 2.1 is

$$n\,\mathrm{ISE}(\hat{F}_n) = n\,\mathrm{ISE}(F_n) + n^{1/2}h^2\mathcal{N} + O(nh^4).$$

Similarly, if $0 < a \leq 1/6$,

$$n\,\mathrm{ISE}(\hat{F}_n) = n\,\mathrm{ISE}(F_n) + n^{1/2}h^2\mathcal{N} + O_p(h).$$

Furthermore, if the kernel is uniform, as pointed out in Remark 1, then the preceding Theorem 2.1 is

$$n\,\mathrm{ISE}(\hat{F}_n) = n\,\mathrm{ISE}(F_n) - 2Ch\int_{-\infty}^{\infty} f^2(x)dx + n^{1/2}h^2\mathcal{N}$$
$$+ O_p(n^{-1/2}h + h^{3/2}) + O(nh^4).$$

Remark 3. In particular, if h is of order $n^{-1/3}$ which minimizing MISE in an asymptotic sense, then we have for the preceding Theorem 2.1,

$$n\,\mathrm{ISE}(\hat{F}_n) = n\,\mathrm{ISE}(F_n) + n^{-1/6}\mathcal{N} + O_p(n^{-1/3}).$$

Remark 4. The useful results by Swanepoel (1988) and others on MISE entail that on the average one can do better with a proper kernel-type estimate than the emprical d.f. In the case of ISE, however, it is impossible to disregard in probability the term which be zero to mean. Note that the term $\mathcal{N}$ of Theorem 2.1 is not independent on $\mathrm{ISE}(F_n)$. From the result of Remark 3, the term of $O_p(n^{-1/6})$ is a random variable following normal distribution, and holds in either case of both positive and negative. Hence there does not exists essentially difference between the estimator by simple empirical d.f. and the estimator which used kernel-type estimate in the sense of ISE.

In the above arguments, the kernel function $k(t)$ is a density function. Now, let us consider another class

(C) $k(t)$ has bounded support and is symmetric around zero such that it satisfies

$$\int_{-\infty}^{\infty} k(t)dt = 1, \qquad \int_{-\infty}^{\infty} tk(t)dt = 0 \quad \text{and} \quad \int_{-\infty}^{\infty} t^2k(t)dt = 0.$$

The condition (C) corresponds to the case $m = 2$ in Falk (1983).

THEOREM 2.2. *Under* (B) *and* (C), *the* ISE *of* $\hat{F}_n(x)$ *is given by*

$$n\,\mathrm{ISE}(\hat{F}_n) = n\,\mathrm{ISE}(F_n) - 2Ch\int_{-\infty}^{\infty} f^2(x)dx + O_p(h) + o_p(h) + o(nh^4),$$

where the expectation of $O_p(h)$ *term is zero.*

PROOF. In this case, $\kappa_2 = 0$. Examine our proofs of Lemmas in Section 3 and the proof of Swanepoel ((1988), Lemmas 4.3–4.6), we then can get the results. Details of the proof are omitted. □

Remark 5. In particular, if $h = O(n^{-1/3})$, then the preceding Theorem 2.2 is

$$n\,\mathrm{ISE}(\hat{F}_n) = n\,\mathrm{ISE}(F_n) - 2Cn^{-1/3}\int_{-\infty}^{\infty} f^2(x)dx + O_p(n^{-1/3}) + o_p(n^{-1/3})$$

where the expectation of the $O_p(n^{-1/3})$ term is zero. For many kernel functions k, the constant C is positive. Thus, though the preceding formula does not necessarily imply the superiority of the kernel estimators, the kernel estimator satisfying (C) will be preferred in the sense of ISE.

3. Proofs of Lemmas

Throughout this section, denote by

$$Y_i(x) = K_h(x - X_i) - L(x), \qquad E(Y_i(x)) = 0,$$
$$H(x_i, x_j) = \int_{-\infty}^{\infty} Y_i(x)Y_j(x)dF(x),$$
$$Z_i = \int_{-\infty}^{\infty} K_h(x - X_i)\{L(x) - F(x)\}dF(x)$$

and

$$A(x) = L(x) - F(x).$$

Then we have

$$I_1 = \frac{1}{n}\sum_{i=1}^{n}\int_{-\infty}^{\infty} Y_i(x)^2 dF(x),$$
$$I_2 = \frac{1}{n}\sum_{i\neq j} H(X_i, X_j), \qquad E\{H(X_i, X_j) \mid X_i\} = 0,$$
$$I_3 = 2\sum_{i=1}^{n}(Z_i - EZ_i),$$
$$I_4 = n\int_{-\infty}^{\infty} A^2(x)dF(x),$$

and by using integration by parts and a Taylor expansion,

$$\begin{aligned} A(x) &= \int_{-\infty}^{\infty} K_h(x - y)dF(y) - F(x) = \int_{-\infty}^{\infty} F(x - ht)k(t)dt - F(x) \\ &= \int_{-\infty}^{\infty}\left\{-htf(x) + \frac{1}{2}h^2t^2 f'(x)\right\}k(t)dt + o(h^2) \\ &= \frac{1}{2}\kappa_2 h^2 f'(x) + o(h^2) \end{aligned}$$

and

$$L(x) = F(x) + \frac{1}{2}\kappa_2 h^2 f'(x) + o(h^2).$$

Hereafter unqualified integral denotes integration over $(-\infty, \infty)$.

PROOF OF LEMMA 2.1. Let $W_i = \int Y_i^2(x)dF(x)$. By using integration by parts and a Taylor expansion we can write,

$$\begin{aligned} W_i &= \int \{K_h(x - X_i) - L(x)\}^2 dF(x) \\ &= \int \{K_h^2(x - X_i) - 2K_h(x - X_i)L(x) + L^2(x)\}dF(x) \end{aligned}$$

$$= 1 - \int F(X_i + ht)2k(t)K(t)dt - \int K_h(x - X_i)\{F^2(x)\}'dx$$
$$+ \frac{1}{3} + O_p(h^2)$$
$$= 1 - F(X_i) - 2Chf(X_i) - \left[1 - F^2(X_i) - 2hF(X_i)f(X_i)\int tk(t)dt\right]$$
$$+ \frac{1}{3} + O_p(h^2)$$
$$= \frac{1}{3} - F(X_i)\{1 - F(X_i)\} - 2Chf(X_i) + O_p(h^2),$$

where $C = \int tk(t)K(t)dt$. Furthermore,

$$E(W_i) = \frac{1}{6} - 2Ch\int f^2(x)dx + o(h),$$
$$E(W_i^2) = E\left[\frac{1}{3} - F(X_i)\{1 - F(X_i)\} - 2Chf(X_i)\right]^2 + o(h),$$
$$= \frac{1}{30} + Ch\int\left[-\frac{4}{3} + 4F(x)\{1 - F(x)\}\right]f(x)dF(x) + o(h),$$
$$E(W_i^3) = \frac{29}{3780} - 6Ch\int\left[F(x)\{1 - F(x)\} - \frac{1}{3}\right]^2 f(x)dF(x) + o(h),$$
$$E(W_i^4) = \frac{11}{5670} - 8Ch\int\left[F(x)\{1 - F(x)\} - \frac{1}{3}\right]^3 f(x)dF(x) + o(h),$$

and so

$$\mathrm{Var}(W_i) = \frac{1}{180} + Ch\int\left[-\frac{2}{3} + 4F(x)\{1 - F(x)\}\right]f(x)dF(x) + o(h),$$
$$E\{W_i - E(W_i)\}^4 = \frac{1}{15120} + O(h).$$

Moreover, put $\sigma_n^2 = \sum_{i=1}^n \mathrm{Var}(W_i)$, then we have

$$\sigma_n^{-4}\sum_{i=1}^n E\{W_i^2 I(|W_i| > \varepsilon\sigma_n^2\} \le \varepsilon^{-2}\sigma_n^{-6}\sum_{i=1}^n E(W_i^4) = O(n^{-1}) \to 0.$$

Hence this completes the proof of Lemma 2.1. □

PROOF OF LEMMA 2.2.

(i) Since

$$\int F(u)K_h(u - x)dF(u)$$
$$= \frac{1}{2} - \frac{1}{2}\int F^2(x + th)k(t)dt$$
$$= \frac{1}{2} - \frac{1}{2}F^2(x) - \frac{1}{2}\kappa_2 h^2\{f^2(x) + F(x)f'(x)\} + o(h^2),$$

and

$$\lim_{h\to 0}\int f'(u)K_h(u-x)dF(u)=\int_x^\infty f'(u)f(u)du=-\frac{1}{2}f^2(x),$$

we have

$$\begin{aligned}&\int L(u)K_h(u-x)dF(u)\\&\quad=\int F(u)K_h(u-x)dF(u)+\frac{1}{2}\kappa_2h^2\int f'(u)K_h(u-x)dF(u)+o(h^2)\\&\quad=\frac{1}{2}-\frac{1}{2}F^2(x)-\frac{1}{2}\kappa_2h^2\left\{\frac{3}{2}f^2(x)+F(x)f'(x)\right\}+o(h^2).\end{aligned}$$

(ii) Similarly,

$$\begin{aligned}\int L^2(u)dF(u)&=\int\{F^2(u)+\kappa_2h^2F(u)f'(u)\}dF(u)\\&=\frac{1}{3}-\frac{1}{2}\kappa_2h^2\int f^3(x)dx+o(h^2).\end{aligned}$$

(iii) Therefore, by (i) and (ii), we have

$$\begin{aligned}J(x,y)=&\int K_h(u-x)K_h(u-y)dF(u)-\int L(u)K_h(u-x)dF(u)\\&-\int L(u)K_h(u-y)dF(u)+\int L^2(u)dF(u)\\&-\left[\frac{1}{3}+\frac{1}{2}\{F^2(x)+F^2(y)\}-F\{\max(x,y)\}\right]\\=&\int K_h(u-x)K_h(u-y)dF(u)-1+F\{\max(x,y)\}\\&+\frac{1}{2}\kappa_2h^2\Big[-\int f^3(x)dx+\frac{3}{2}\{f^2(x)+f^2(y)\}\\&\qquad+F(x)f'(x)+F(y)f'(y)\Big]+o(h^2).\end{aligned}$$

(iv) The problem is to evaluate $\int K_h(u-x)K_h(u-y)dF(u)$. Without loss of generality we may assume that $k(t)=0$ for $|t|>1/2$, i.e., $f(x)$ is continuous in $(x-1/2,x+1/2)$, and $x>y$. At first, we consider the case of $x-y>h$, then

$$\begin{aligned}&\int K_h(u-x)K_h(u-y)dF(u)\\&\quad=\int_{u>x+h/2}dF(u)+\int_{x-h/2}^{x+h/2}K_h(u-x)f(u)du\\&\quad=1-F\left(x+\frac{h}{2}\right)+h\int_{-1/2}^{1/2}K(t)f(x+ht)dt\end{aligned}$$

$$\begin{aligned}
&= 1 - F(x) - \frac{h}{2}f(x) - \frac{h^2}{8}f'(x) \\
&\quad + h\int_{-1/2}^{1/2} K(t)\{f(x) + htf'(x)\}dt + o(h^2) \\
&= 1 - F(x) + hf(x)\left\{-\frac{1}{2} + \int_{-1/2}^{1/2} K(t)dt\right\} \\
&\quad + h^2 f'(x)\left\{-\frac{1}{8} + \int_{-1/2}^{1/2} tK(t)dt\right\} + o(h^2).
\end{aligned}$$

Second, in the case of $0 < x - y < h$, we have

$$\begin{aligned}
&\int K_h(u-x)K_h(u-y)dF(u) \\
&\quad = \int_{u>x+h/2} dF(u) + \int_{y+h/2}^{x+h/2} K_h(u-x)f(u)du \\
&\qquad + \int_{x-h/2}^{y+h/2} K_h(u-x)K_h(u-y)f(u)du \\
&\quad = 1 - F\left(x + \frac{h}{2}\right) + h\int_{1/2-z}^{1/2} K(t)f(x+ht)dt \\
&\qquad + h\int_{-1/2}^{1/2-z} K(t)K(z+t)f(x+ht)dt \\
&\quad = 1 - F(x) - \frac{h}{2}f(x) - \frac{h^2}{8}f'(x) + h\int_{1/2-z}^{1/2} K(t)\{f(x) + htf'(x)\}dt \\
&\qquad + h\int_{-1/2}^{1/2-z} K(t)K(z+t)\{f(x) + htf'(x)\}dt + o(h^2) \\
&\quad = 1 - F(x) + hf(x)\left[-\frac{1}{2} + \int_{1/2-z}^{1/2} K(t)dt + \int_{-1/2}^{1/2-z} K(t)K(z+t)dt\right] \\
&\qquad + h^2 f'(x)\left[-\frac{1}{8} + \int_{1/2-z}^{1/2} tK(t)dt + \int_{-1/2}^{1/2-z} tK(t)K(z+t)dt\right] + o(h^2),
\end{aligned}$$

where $z = (x-y)/h$. We can not proceed the above arguments for kernel functions with non-compact support. Thus, we need the compactness of the support of $k(t)$ to prove (iv) in the proof of this lemma.

In order to end the proof, we may consider the following statement. Let the eigenvalues and eigenfunctions of $J(x, y)$ be (λ_1, p_1), $(\lambda_2, p_2), \dots$, i.e., the solutions of

$$\int J(x, y)p(y)dF(y) = \lambda p(x),$$

satisfying

$$\int p_i(x)dF(x) = 0, \quad \int p_i^2(x)dF(x) = 1 \quad \text{and} \quad \int_{i \neq j} p_i(x)p_j(x)dF(x) = 0.$$

Then we have

$$J(x,y)=\sum_{k=1}^{\infty}\lambda_k p_k(x)p_k(y),$$
$$I_{21}=\sum_{k=1}^{\infty}\lambda_k\left[\left\{n^{-1/2}\sum_{i=1}^{n}p_k(X_i)\right\}^2-\frac{1}{n}\sum_{i=1}^{n}p_k^2(X_i)\right],$$
$$E(I_{21}^2)=n^{-2}\sum_{k,i\neq j,i'\neq j'}\lambda_k^2E\{p_k(X_i)p_k(X_j)p_k(X_{i'})p_k(X_{j'})\}$$
$$=2(n-1)n^{-1}\sum_{k=1}^{\infty}\lambda_k^2$$

and

$$\sum_{k=1}^{\infty}\lambda_k^2=\int J^2(x,y)dF(x)dF(y).$$

From the preceding results and (iii)–(iv), we have

$$\int J^2(x,y)dF(x)F(y)=2\int_{x>y}J^2(x,y)dF(x)F(y)=O(h^2),$$

since $J(x,y)$ is symmetric about x, y. Hence,

$$E(I_{21}^2)=O(h^2)\quad\text{and}\quad Pr(h^{-1}|I_{21}|>\varepsilon)\le\varepsilon^{-2}h^{-2}E(I_{21}^2).$$

The proof of Lemma 2.2 is completed. □

PROOF OF LEMMA 2.3. Since $Z_1,\ldots,Z_n$ are independently and identically distributed,

$$E(Z_i)=\int L(x)\{L(x)-F(x)\}dF(x)=\int A(x)\{A(x)+F(x)\}dF(x)$$
$$=\frac{\kappa_2^2h^4}{4}\int\{f'(x)\}^2dF(x)+\frac{\kappa_2h^2}{2}\int f'(x)F(x)dF(x)+o(h^2)$$
$$=-\frac{\kappa_2h^2}{4}\int f^3(x)dx+o(h^2)$$

and

$$E(Z_i^2)=\int\left\{\int K_h(x-u)A(x)dF(x)\right\}^2dF(u)$$
$$=\int\left[\int\frac{\kappa_2h^2}{4}K_h(x-u)\{f^2(x)\}'dx\right]^2dF(u)+o(h^4)$$
$$=\frac{\kappa_2^2h^4}{16}\int\left\{\int f^2(x+ht)k(t)dt\right\}^2dF(x)+o(h^4)$$
$$=\frac{\kappa_2^2h^4}{16}\int f^5(x)dx+o(h^4).$$

The desired results follow immediately, on noting that

$$\mathrm{Var}(Z_i) = \frac{\kappa_2^2 h^4}{16} \mathrm{Var}\{f^2(X)\} + o(h^4) \quad \text{and} \quad E(Z_i^4) = O(h^8). \qquad \square$$

Acknowledgements

The authors would like to express their sincere thanks to the referees for their very careful advice and many useful comments.

References

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

Falk, M. (1983). Relative efficiency and deficiency of kernel type estimators of smooth distribution functions, *Statist. Neerlandica*, **37**, 73–83.

Hall, P. (1984). Central limit theorem for integrated square error of multivariate nonparametric density estimators, *J. Multivariate Anal.*, **14**, 1–16.

Hill, P. D. (1985). Kernel estimation of a distribution function, *Comm. Statist. Theory Methods*, **14**, 605–620.

Jones, M. C. (1990). The performance of kernel density functions in kernel distribution function estimation, *Statist. Probab. Lett.*, **9**, 129–132.

Mack, Y. P. (1984). Remarks on some smoothed empirical distribution functions and processes, *Bull. Inform. Cybernet.*, **21**, 29–35.

Mammitzsch, V. (1984). On the asymptotically optimal solution within a certain class of kernel type estimatiors, *Statist. Decisions*, **2**, 247–255.

Nadaraya, E. A. (1964). Some new estimates for distribution function, *Theory Probab. Appl.*, **9**, 497–500.

Parzen, E. (1962). On the estimation of a probability density and mode, *Ann. Math. Statist.*, **33**, 1065–1076.

Reiss, R. D. (1981). Nonparametric estimation of smooth distribution functions, *Scand. J. Statist.*, **8**, 116–119.

Rosenblatt, M. (1956). Remarks on some non-parametric estimates of a density function, *Ann. Math. Statist.*, **27**, 832–837.

Swanepoel, J. W. H. (1988). Mean integrated squared error properties and optimal kernels when estimating a distribution function, *Comm. Statist. Theory Methods*, **17**, 3785–3799.

Watson, G. S. and Leadbetter, M. R. (1964). Hazard analysis II, *Sankhyā Ser. A*, **26**, 101–116.

Winter, B. B. (1973). Strong uniform consistency of integrals of density estimators, *Canad. J. Statist.*, **1**, 247–253.

Winter, B. B. (1979). Convergence rate of perturbed empirical distribution functions, *J. Appl. Probab.*, **16**, 163–173.

Yamato, H. (1973). Uniform convergence of an estimator of a distribution function, *Bull. Math. Statist.*, **15**, 69–78.