T E C H N I C A L
R E P O R T

0458

MODIFYING THE KERNEL DISTRIBUTION ESTIMATOR TOWARDS REDUCED BIAS

Paul JANSSEN, Jan SWANEPOEL and Noël VERAVERBEKE

I A P S T A T I S T I C S
N E T W O R K

INTERUNIVERSITY ATTRACTION POLE

http://www.stat.ucl.ac.be/IAP

MODIFYING THE KERNEL DISTRIBUTION FUNCTION ESTIMATOR TOWARDS REDUCED BIAS

Paul JANSSEN[1], Jan SWANEPOEL[2], Noël VERAVERBEKE[1]

[1] Limburgs Universitair Centrum, Diepenbeek, Belgium
[2] North-West University, Potchefstroom Campus, South Africa

SUMMARY

We explore the convergence rates of a kernel-based distribution function estimator with variable bandwidth. As in density estimation, a considerable bias reduction from $O(h^2)$ to $O(h^4)$ can be obtained by replacing the bandwidth h by $h/f^{1/2}(X_i)$. We show that the necessary replacement of $f^{1/2}$ by some pilot estimator $\widehat{f}_g^{\,1/2}$, depending on a second bandwidth g, has no penalizing effect on bias and variance, provided we undersmooth with the pilot bandwidth g, that is $g/h \to 0$ in a certain way. Due to the considerable bias reduction a simple plug-in normal reference bandwidth selector works effectively in practice. Distribution function estimators with good convergence properties and with simple bandwidth selectors are desirable for repetitive use in smoothed bootstrap algorithms.

Some key words: Bandwidth; Bias; Distribution function estimator; Kernel estimation; Mean integrated squared error; Smoothed bootstrap

AMS Subject classification: MSC2000; Primary 62N02, Secondary 62F40

1. INTRODUCTION

If $X_1, \ldots, X_n$ is a random sample from a smooth distribution function (df) F, then the classical nonparametric kernel estimator for F is given by

$$\widehat{F}_{n,h}(x) = \frac{1}{n} \sum_{i=1}^{n} K\left(\frac{x - X_i}{h}\right), \tag{1.1}$$

where K is a known df having a density k that is symmetric around zero and where h is the bandwidth. We will assume throughout that k has unbounded support $(-\infty, +\infty)$. Under the conditions that F has two continuous derivatives f and f', it is well known (see e.g. Azzalini (1981)) that, as $n \to \infty$,

$$E(\widehat{F}_{n,h}(x)) = F(x) + \frac{1}{2} h^2 f'(x) \mu_2(k) + o(h^2), \tag{1.2}$$

$$\mathrm{Var}(\widehat{F}_{n,h}(x)) = \frac{F(x)(1 - F(x))}{n} - \frac{2hf(x)c_1}{n} + o\left(\frac{h}{n}\right), \tag{1.3}$$

where $\mu_\ell(k) = \int_{-\infty}^{+\infty} z^\ell k(z)dz$ $(\ell = 2, 4)$ and $c_1 = \int_{-\infty}^{+\infty} zk(z)K(z)dz$.
Optimal choices for the bandwidth h can be obtained by minimizing the approximate mean integrated squared error (MISE) calculated from the expressions in (1.2) and (1.3).

There have been many proposals in the literature for improving the bias property of the basic kernel density estimator. Typically, under sufficient smoothness conditions placed on the underlying density f, the bias reduces from $O(h^2)$ to $O(h^4)$, and the variance remains of order $(nh)^{-1}$. Those density estimators, that could potentially have greater practical impact, include variable bandwidth kernel estimators (Abramson, 1982), variable location estimators (Samiuddin & El-Sayyad, 1990), nonparametric transformation estimators (Ruppert & Cline, 1994) and multiplicative bias correction estimators (Jones *et al.*, 1995). A comparative study was carried out by Jones & Signorini (1997) for all these estimators with bias $O(h^4)$. However it seems that hardly any work has been done to reduce the bias of the classical df estimator $\widehat{F}_{n,h}$ in (1.1). Nevertheless, an important motivation for designing a 'better' df estimator is its use in the smoothed bootstrap. The accuracy of the latter clearly depends on the quality of the proposed smooth estimator for F. For example, one could use a so-called higher-order kernel K, but this has the disadvantage of not leading to bona fide df estimates. Another approach, which is based on a nonparametric transformation of the data, has recently been developed in Swanepoel & Van Graan (2005). They propose to use an estimator

$$\widetilde{F}_n(x) = \frac{1}{n}\sum_{i=1}^{n} K\left(\frac{\widehat{F}_{n,g}(x) - \widehat{F}_{n,g}(X_i)}{h}\right), \tag{1.4}$$

where $\widehat{F}_{n,g}$ is the kernel estimator in (1.1) with bandwidth g. They obtain a serious bias reduction and propose an easy method for choosing the bandwidths g and h.

In the present paper we study a different estimator which is obtained by considering a variable bandwidth version of (1.1). It consists of replacing the bandwidth h by $h/\widehat{f_g}^{1/2}(X_i)$, where $\widehat{f_g}$ is some pilot estimator for the density function f. Our df estimator is defined as

$$\widehat{F}_{n,h,g}(x) = \frac{1}{n}\sum_{i=1}^{n} K\left(\frac{x - X_i}{h}\widehat{f_g}^{1/2}(X_i)\right), \tag{1.5}$$

where

$$\widehat{f_g}(x) = \frac{1}{ng}\sum_{i=1}^{n} k\left(\frac{x - X_i}{g}\right), \tag{1.6}$$

is the Parzen-Rosenblatt kernel density estimator for $f = F'$ with kernel $k = K'$ and bandwidth g. The form of the adaptation in (1.5) is found after a careful study of the asymptotic MISE in Section 2. We prove that the bias reduces from $O(h^2)$ to $O(h^4)$ and that the bias reduction has no penalizing effect on the $O(h/n)$ term in the variance, keeping the same order compared to the classical kernel estimator, but with a different constant. It requires that we undersmooth with the pilot bandwidth g in the sense that $g/h \to 0$ and $h^3/g \to 0$. In Section 3, we present a plug-in bandwidth selector that can be calculated in an easy and quick way. This is desirable in view of its repetitive use in the generation of smoothed bootstrap samples. We give the algorithm in Section 4. Proofs are given in Section 5.

Two different ways of bias reduction in df estimation (variable location method and multiplicative method) will be the content of a forthcoming paper by the same authors.

2. Asymptotic expansion for mean and variance

Our starting point is to take expression (1.5), but we replace $\widehat{f_g}^{1/2}$ temporarily by some general smooth function φ. We then put

$$T_n = \frac{1}{n}\sum_{i=1}^{n} K\left(\frac{x - X_i}{h}\varphi(X_i)\right). \tag{2.1}$$

The function φ will be specified after we have calculated expressions for bias and variance of T_n. This is done in the following theorem.

Theorem 1. Assume

(C1) F has four continuous derivatives in a neighborhood of x and $f(x) = F'(x) > 0$.

(C2) K is a df with unbounded support $(-\infty, +\infty)$, having a density k that is symmetric around zero. The density k is bounded and has four continuous derivatives $k^{(i)}$ $(i = 1, \dots, 4)$ and $|u|^6 k(u)$, $|u|^7 |k^{(1)}(u)|$, $|u|^8 |k^{(2)}(u)|$ and $|u|^8 |k^{(3)}(u)|$ all tend to zero as u tends to $+\infty$ or $-\infty$.

Suppose that h is a bandwidth sequence tending to zero as $n \to \infty$. Then, for any function φ, having three continuous derivatives in a neighborhood of x and $\varphi(x) > 0$, we have, as $n \to \infty$,

$$E(T_n) = F(x) + h^2 Q_0(x) + h^4 Q_1(x) + o(h^4), \tag{2.2}$$

$$\mathrm{Var}(T_n) = \frac{F(x)(1 - F(x))}{n} - 2\frac{h}{n}\frac{f(x)}{\varphi(x)} c_1 + o\left(\frac{h}{n}\right), \tag{2.3}$$

where

$$Q_0(x) = \frac{1}{2}\frac{\mu_2(k)}{\varphi^3(x)}\left[f'(x)\varphi(x) - 2f(x)\varphi'(x)\right],$$

$$\begin{aligned} Q_1(x) = \frac{\mu_4(k)}{24\varphi^7(x)} \big[&-120 f(x)(\varphi'(x))^3 + 60 f(x)\varphi(x)\varphi'(x)\varphi''(x) \\ &-4 f(x)\varphi^2(x)\varphi'''(x) + 60 f'(x)\varphi(x)(\varphi'(x))^2 \\ &-12 f'(x)\varphi^2(x)\varphi''(x) + 12 f''(x)\varphi^2(x)\varphi'(x) \\ &+ f'''(x)\varphi^3(x)\big]. \end{aligned}$$

The proof of this theorem is a straightforward but lengthy calculation. A sketch is given in Section 5.

To obtain an estimator with reduced bias, we choose the function φ in such a way that $Q_0(x)$ vanishes, i.e. φ is such that, for all x, it satisfies the differential equation

$$f'(x)\varphi(x) - 2f(x)\varphi'(x) = 0.$$

This leads to the choice $\varphi(x) = f^{1/2}(x)$, and in this case

$$E(T_n) = F(x) + h^4 \widetilde{Q}_1(x) + o(h^4), \tag{2.4}$$

where

$$\widetilde{Q}_1(x) = \frac{\mu_4(k)}{24 f^4(x)} \left[6 f(x) f'(x) f''(x) - 6(f'(x))^3 - f^2(x) f'''(x) \right]. \quad (2.5)$$

Our estimator in (1.5) is then obtained by replacing φ in (2.1) by $\widehat{f_g}^{1/2}$, where $\widehat{f_g}$ is the estimator for f given in (1.6).
The next theorem summarizes the effect of this substitution and it provides the mean and variance of the estimator $\widehat{F}_{n,h,g}$ in (1.5).

Theorem 2. Assume (C1) and (C2) of Theorem 1. Suppose that g and h are bandwidths tending to zero as $n \to \infty$ and such that $g/h \to c$ as $n \to \infty$, for some constant c, $0 \le c < \infty$. Then it follows that

(a) if $ngh^3 \to \infty$ and $ng^2 \to \infty$ as $n \to \infty$,

$$\begin{aligned} E\left(\widehat{F}_{n,h,g}(x)\right) &= F(x) + h^4\, \widetilde{Q}_1(x) \\ &+ \frac{1}{4} g^2 h^2 \mu_2^2(k) \left[\frac{f'(x) f''(x)}{f^2(x)} - \frac{f'''(x)}{f(x)} \right] + o(h^4), \end{aligned} \quad (2.6)$$

where $\widetilde{Q}_1(x)$ is given in (2.5).

(b) if $ngh^3 \to \infty$ and $ngh^3/ng^2 = h^3/g \to 0$ as $n \to \infty$,

$$\mathrm{Var}\left(\widehat{F}_{n,h,g}(x)\right) = \frac{F(x)(1-F(x))}{n} - 2\frac{h}{n} f^{1/2}(x) c_1 + o\left(\frac{h}{n}\right). \quad (2.7)$$

The proof of this theorem is sketched in Section 5.

Note that the first term in the variance expression remains unchanged compared to that of the kernel df estimator, and that the second term is also of the order $O(h/n)$. Also note that the choice $g/h \to c$ with $c > 0$ leads to an additional term of order $O(h^4)$ in the bias expression. Therefore we recommend to take a pilot bandwidth g for which $g/h \to 0$. This gives a bias expression

$$E(\widehat{F}_{n,h,g}(x)) = F(x) + h^4 \widetilde{Q}_1(x) + o(h^4). \quad (2.8)$$

Remark 1.
Note that for the coefficient $\widetilde{Q}_1(x)$ of h^4 we also have that

$$\begin{aligned}\widetilde{Q}_1'(x) &= \frac{\mu_4(k)}{24}\{-36f^{-4}(x)(f'(x))^2f''(x)+8f^{-3}(x)f'(x)f'''(x)\\&+6f^{-3}(x)(f''(x))^2-f^{-2}(x)f^{(4)}(x)+24f^{-5}(x)(f'(x))^4\}\end{aligned}$$

in correspondence with the density counterpart in e.g. Silverman (1986), and also that

$$\widetilde{Q}_1(x)=\frac{\mu_4(k)}{24}\left(\frac{1}{f(x)}\right)'''$$

in correspondence with the integral of the formula in Hall, Hu & Marron (1995).

The MISE is a typical global measure of accuracy and we therefore consider the asymptotic expression for

$$MISE\left(\widehat{F}_{n,h,g}\right)=E\left[\int_{-\infty}^{+\infty}(\widehat{F}_{n,h,g}(x)-F(x))^2w(x)dF(x)\right], \tag{2.9}$$

where w is some weight function.
Under the conditions of Theorem 2 and with $g/h\to 0$, we obtain that, as $n\to\infty$,

$$MISE\left(\widehat{F}_{n,h,g}\right)\cong\frac{1}{n}\int_{-\infty}^{+\infty}F(x)(1-F(x))w(x)dF(x)$$

$$-2\frac{h}{n}c_1\int_{-\infty}^{+\infty}f^{1/2}(x)w(x)dF(x)+h^8\int_{-\infty}^{+\infty}\widetilde{Q}_1^2(x)w(x)dF(x),$$

with $\widetilde{Q}_1(x)$ as in (2.5). From this it follows that the asymptotic optimal bandwidth is

$$h_{opt}=\left(\frac{c_1\int_{-\infty}^{+\infty}f^{1/2}(x)w(x)dF(x)}{4\int_{-\infty}^{+\infty}\widetilde{Q}_1^2(x)w(x)dF(x)}\right)^{1/7}n^{-1/7}, \tag{2.10}$$

and that for this optimal bandwidth the approximate MISE is given by

$$\frac{1}{n}\int_{-\infty}^{+\infty}F(x)(1-F(x))w(x)dF(x)-\frac{7}{4}\frac{\left(c_1\int_{-\infty}^{+\infty}f^{1/2}(x)w(x)dF(x)\right)^{8/7}}{\left(4\int_{-\infty}^{+\infty}\widetilde{Q}_1^2(x)w(x)dF(x)\right)^{1/7}}n^{-8/7}.$$

The minus sign of the second term in this expression clearly shows the improvement compared to the classical empirical distribution function. The order of the improvement $n^{-8/7}$ also compares favorably to that of the classical kernel estimator, *viz.* $n^{-4/3}$ which can be derived from (1.2) and (1.3), and to that of the estimator of Swanepoel and Van Graan (2005), which is $n^{-16/15-\delta}$, for arbitrary $\delta > 0$.

3. BANDWIDTH SELECTION

Because the estimator has small bias, we propose to use a simple plug-in normal reference bandwidth selector (see e.g. Silverman (1986)). We choose the weight function w in (2.9) as $w(x) = \left(\frac{1}{\sigma}\phi\left(\frac{x}{\sigma}\right)\right)^2$ and calculate h_{opt} in (2.10) explicitly for $f(x) = \frac{1}{\sigma}\phi\left(\frac{x}{\sigma}\right)$, where ϕ is the standard normal density function and σ the population standard deviation. Substituting $f'(x) = -\frac{x}{\sigma^3}\phi\left(\frac{x}{\sigma}\right)$, $f''(x) = \frac{1}{\sigma^3}\left(\frac{x^2}{\sigma^2} - 1\right)\phi\left(\frac{x}{\sigma}\right)$ and $f'''(x) = \frac{1}{\sigma^4}\left(\frac{3x}{\sigma} - \frac{x^3}{\sigma^3}\right)\phi\left(\frac{x}{\sigma}\right)$ into the definition of $\widetilde{Q}_1$ in (2.5), we obtain

$$\widetilde{Q}_1(x) = \frac{\mu_4(k)}{24\phi\left(\frac{x}{\sigma}\right)}\left[\frac{3x}{\sigma^3} + \frac{x^3}{\sigma^5}\right].$$

This gives

$$\int_{-\infty}^{+\infty} \widetilde{Q}_1^2(x)w(x)dF(x) = \frac{7\mu_4^2(k)}{96\sigma^6}.$$

Also

$$\int_{-\infty}^{+\infty} f^{1/2}(x)w(x)dF(x) = \frac{1}{\sigma^{5/2}}\sqrt{\frac{2}{7}}\frac{1}{(2\pi)^{5/4}}.$$

From (2.10), it follows that

$$h_{opt} = \left(\frac{24c_1}{7\mu_4^2(k)(2\pi)^{5/4}}\left(\frac{2}{7}\right)^{1/2}\right)^{1/7}\sigma^{1/2}n^{-1/7}.$$

This gives a bandwidth selector $\widehat{h}$ if in this formula we replace σ by the sample standard deviation S, or by a measure suggested by Silverman (1986, p.47) that is more robust to deviations from normality, which is given by $\widehat{\sigma} = \min(S, IQR/1.349)$, where IQR is the interquartile range.

In the example of a standard normal kernel $K = \Phi$, we have $\mu_2(k) = 1$, $\mu_4(k) = 3$, $c_1 = \dfrac{1}{2\sqrt{\pi}}$ and

$$\widehat{h} = 0.479\ \widehat{\sigma}^{1/2} n^{-1/7}.$$

Here $\widehat{g}$ can e.g. be taken as $\widehat{g} = \widehat{c}\ \widehat{h}^{\alpha}$ for some constants $0 < \widehat{c} < \infty$ and $\alpha > 1$.

Remark 2.
The expression for the bandwidth selector is somewhat more complicated if we work under the conditions of Theorem 2 with $g/h \to c > 0$. A similar calculation with $\widetilde{Q}_1(x)$ replaced by

$$\frac{\mu_4(k)}{24\phi\left(\frac{x}{\sigma}\right)}\left[\frac{3x}{\sigma^3} + \frac{x^3}{\sigma^5}\right] - c^2\frac{\mu_2^2(k)}{2\sigma^4}x$$

gives that in this case

$$h_{opt} = \left(\frac{c_1}{4} \frac{\frac{1}{\sigma^{5/2}}\sqrt{\frac{2}{7}}\frac{1}{(2\pi)^{5/4}}}{\frac{7\mu_4^2(k)}{96\sigma^6} + c^4\frac{\mu_2^4(k)}{24\sqrt{3}\pi\sigma^8} - c^2\frac{3\mu_4(k)\mu_2^2(k)}{64\sqrt{\pi}\sigma^7}} \right)^{1/7} n^{-1/7}.$$

In the example of a standard normal kernel $K = \Phi$, we have

$$\widehat{h} = 0.685\left(12.213\widehat{\sigma}^{-7/2} + 0.144c^4\widehat{\sigma}^{-11/2} - 1.488c^2\widehat{\sigma}^{-9/2}\right)^{-1/7} n^{-1/7},$$

and $\widehat{g} = c\widehat{h}$. It is easily verified that the quantity in brackets is strictly positive. Preliminary Monte Carlo results show that this $\widehat{h}$ and $\widehat{g}$ work effectively in practice.

4. ILLUSTRATION: GENERATING BOOTSTRAP SAMPLES

Generating bootstrap samples requires a good quality estimator for F. For anyone who wants to perform bootstrap calculations, the following algorithm based on the estimator $\widehat{F}_{n,h,g}$ can be used:

(1) obtain $X_1^*, \dots, X_n^*$ by sampling with replacement from $X_1, \dots, X_n$

(2) generate, independently, a random sample $Z_1, \dots, Z_n$ from K

(3) put $Y_i^* = X_i^* + hZ_i/\widehat{f_g}^{1/2}(X_i^*)$.

Step (3) in the algorithm clearly shows how the variable bandwidth adapts to the available information. Our estimator is easier to use than the estimator $\widetilde{F}_n(x)$ in (1.4). There $\widetilde{Y}_i^* = \widetilde{F}_n^{-1}(\widetilde{X}_i^* + hZ_i)$, where $\widetilde{X}_1^*, \ldots, \widetilde{X}_n^*$ are obtained by sampling with replacement from $\widehat{F}_{n,g}(X_i)$, $i = 1, \ldots, n$. Taking the inverse to generate $\widetilde{Y}_i^*$ is problematic if the argument falls outside the range $[0, 1]$.

5. PROOFS

5.1 Proof of Theorem 1

Integration by parts gives that

$$E(T_n) = \int_{-\infty}^{+\infty} F(x - hz)k(zg(x - hz))[g(x - hz) - hzg'(x - hz)]dz.$$

Next we expand $F(x - hz)$, $k(zg(x - hz))$, $g(x - hz)$ and $g'(x - hz)$ in a Taylor series up to terms with $(hz)^4$. Then follows a long and tedious calculation where we also use that $\int_{-\infty}^{+\infty} u^m k'(u)du = 0$ if m is even, $= -3\mu_2(k)$ if $m = 3$, $= -5\mu_4(k)$ if $m = 5$; $\int_{-\infty}^{+\infty} u^m k''(u)du = 0$ if m is odd, $= 12\mu_2(k)$ if $m = 4$, $= 30\mu_4(k)$ if $m = 6$; $\int_{-\infty}^{+\infty} u^m k'''(u)du = 0$ if m is even, $= -210\mu_4(k)$ if $m = 7$ and $\int_{-\infty}^{+\infty} u^8 k^{(4)}(u)du = 1680\mu_4(k)$. Collecting all the terms leads to the expansion of $E(T_n)$ as given in Theorem 1. A similar calculation is done for the variance:

$$\begin{aligned} \mathrm{Var}(T_n) &= \frac{1}{n}\mathrm{Var}\left(K\left(\frac{x - X_1}{h}g(X_1)\right)\right) \\ &= \frac{1}{n}\left[E\left(K^2\left(\frac{x - X_1}{h}g(X_1)\right)\right) - F^2(x) + O(h^2)\right] \end{aligned}$$

and

$$\begin{aligned} &E\left(K^2\left(\frac{x - X_1}{h}g(X_1)\right)\right) \\ &= 2\int_{-\infty}^{+\infty} F(x - hz)K(zg(x - hz))k(zg(x - hz)) \\ &\qquad [g(x - hz) - hzg'(x - hz)]dz. \end{aligned}$$

5.2 Proof of Theorem 2

Applying the Taylor approximations

$$\frac{1}{n}\sum_{i=1}^{n} K\left(\frac{x-X_i}{h}\widehat{f}_g^{\,1/2}(X_i)\right) \cong \frac{1}{n}\sum_{i=1}^{n} K\left(\frac{x-X_i}{h}f^{1/2}(X_i)\right)$$
$$+\frac{1}{n}\sum_{i=1}^{n} k\left(\frac{x-X_i}{h}f^{1/2}(X_i)\right)\left(\frac{x-X_i}{h}\right)\left(\widehat{f}_g^{\,1/2}(X_i)-f^{1/2}(X_i)\right),$$

and also

$$\widehat{f}_g^{\,1/2}(X_i)-f^{1/2}(X_i) \cong \frac{1}{2}\frac{\widehat{f}_g(X_i)-f(X_i)}{f^{1/2}(X_i)}$$
$$\cong \frac{1}{2}\left[\frac{\widehat{f}_g(x)-f(x)}{f^{1/2}(x)}+\frac{\widehat{f}_g'(x)-f'(x)}{f^{1/2}(x)}(X_i-x)-\frac{\widehat{f}_g(x)-f(x)}{2f^{3/2}(x)}f'(x)(X_i-x)\right],$$

we obtain

$$\widehat{F}_{n,h,g}(x) \cong S_n+\Sigma_1+\Sigma_2+\Sigma_3, \tag{5.1}$$

where

$$S_n = \frac{1}{n}\sum_{i=1}^{n} K\left(\frac{x-X_i}{h}f^{1/2}(X_i)\right),$$
$$\Sigma_1 = \frac{1}{2n}\sum_{i=1}^{n} k\left(\frac{x-X_i}{h}f^{1/2}(X_i)\right)\left(\frac{x-X_i}{h}\right)\frac{\widehat{f}_g(x)-f(x)}{f^{1/2}(x)},$$
$$\Sigma_2 = \frac{1}{2n}\sum_{i=1}^{n} k\left(\frac{x-X_i}{h}f^{1/2}(X_i)\right)\left(\frac{x-X_i}{h}\right)\frac{\widehat{f}_g'(x)-f'(x)}{f^{1/2}(x)}(X_i-x),$$
$$\Sigma_3 = -\frac{1}{2n}\sum_{i=1}^{n} k\left(\frac{x-X_i}{h}f^{1/2}(X_i)\right)\left(\frac{x-X_i}{h}\right)\frac{\widehat{f}_g(x)-f(x)}{2f^{3/2}(x)}f'(x)(X_i-x).$$

The expectation of the first term in the right hand side of (5.1) is given in (2.5).
For Σ_1 we have , with $A_i = k\left(\frac{x-X_i}{g}\right) - gf(x)$ and $B_i = k\left(\frac{x-X_i}{h}f^{1/2}(X_i)\right)\left(\frac{x-X_i}{h}\right)$,

$i = 1, \ldots, n$, that

$$\begin{aligned} E(\Sigma_1) &= \frac{1}{2f^{1/2}(x)n^2 g} E\left(\sum_{i=1}^n A_i \sum_{j=1}^n B_j\right) \\ &= \frac{1}{2f^{1/2}(x)n^2 g}\left(\sum_{i=1}^n E(A_i B_i) + \sum_{i=1}^n \sum_{\substack{j=1 \\ j\neq i}}^n E(A_i)E(B_j) \right) \\ &= \frac{1}{2f^{1/2}(x)n^2 g}\left(nE(A_1B_1) + n(n-1)E(A_1)E(B_1)\right). \end{aligned}$$

Applying Taylor series expansions, we find that $E(A_1) = \frac{1}{2}g^3 f''(x)\mu_2(k) + o(g^3)$, $E(B_1) = \frac{1}{2}h^2 \frac{f'(x)}{f^{3/2}(x)}\mu_2(k) + o(h^2)$ and $E(A_1B_1) = O(g^2/h)$ if $c > 0$ and $o(g^2/h)$ if $c = 0$.

And hence, under the assumptions that $ngh^3 \to \infty$ and $ng^2 \to \infty$, it readily follows that

$$E(\Sigma_1) = \frac{1}{8}g^2h^2\frac{f'(x)f''(x)}{f^2(x)}\mu_2^2(k) + o(g^2h^2).$$

Completely similar we find that

$$\begin{aligned} E(\Sigma_2) &= -\frac{1}{4}g^2h^2\frac{f'''(x)}{f(x)}\mu_2^2(k) + o(g^2h^2) \\ E(\Sigma_3) &= \frac{1}{8}g^2h^2\frac{f'(x)f''(x)}{f^2(x)}\mu_2^2(k) + o(g^2h^2). \end{aligned}$$

Combining the above we find the expansion for $E(\widehat{F}_{n,h,g}(x))$.

A similar treatment holds for $\mathrm{Var}(\widehat{F}_{n,h,g}(x))$. From (5.1) we have that

$$\begin{aligned} \mathrm{Var}(\widehat{F}_{n,h,g}(x)) \cong\ & \mathrm{Var}(S_n) + \mathrm{Var}(\Sigma_1) + \mathrm{Var}(\Sigma_2) + \mathrm{Var}(\Sigma_3) \\ & +2(\mathrm{Cov}(S_n, \Sigma_1) + \mathrm{Cov}(S_n, \Sigma_2) + \mathrm{Cov}(S_n, \Sigma_3) + \mathrm{Cov}(\Sigma_1, \Sigma_2) \\ & +\mathrm{Cov}(\Sigma_2, \Sigma_3) + \mathrm{Cov}(\Sigma_1, \Sigma_3)). \end{aligned}$$

The variance of S_n is given in (2.3) with $\varphi = f^{1/2}$. For Σ_1 we have

$$\begin{aligned} \mathrm{Var}(\Sigma_1) &= \frac{1}{4f(x)n^4g^2}\mathrm{Var}\left(\sum_{i=1}^n A_i \sum_{j=1}^n B_j\right) \\ &= \frac{1}{4f(x)n^4g^2}\left[E\left\{\left(\sum_{i=1}^n A_i\right)^2\left(\sum_{j=1}^n B_j\right)^2\right\} - \left\{E\left[\left(\sum_{i=1}^n A_i\right)\left(\sum_{j=1}^n B_j\right)\right]\right\}^2\right]. \end{aligned}$$

A careful examination of all possible terms shows that under the stated conditons

$$\begin{aligned}\mathrm{Var}(\Sigma_1) &= \frac{1}{4f(x)n^4g^2}\left\{n(n-1)(n-2)(n-3)(E(A_1)E(B_1))^2\right.\\ &\quad \left.-n^2(n-1)^2(E(A_1)E(B_1))^2\right\} + o\left(\frac{h}{n}\right)\\ &= o\left(\frac{h}{n}\right).\end{aligned}$$

Completely similar we find that $\mathrm{Var}(\Sigma_2)$, $\mathrm{Var}(\Sigma_3)$, $\mathrm{Cov}(S_n, \Sigma_1)$, $\mathrm{Cov}(S_n, \Sigma_2)$, $\mathrm{Cov}(S_n, \Sigma_3)$ are all $o\left(\frac{h}{n}\right)$. Also $\mathrm{Cov}(\Sigma_1, \Sigma_2)$, $\mathrm{Cov}(\Sigma_2, \Sigma_3)$ and $\mathrm{Cov}(\Sigma_1, \Sigma_3)$ are $o\left(\frac{h}{n}\right)$ by a simple application of the Cauchy-Schwarz inequality.

Acknowledgement

The Belgian Federal Science Policy Office provided partial support for Janssen and Veraverbeke (Interuniversity Attraction Pole Programme P5/24). Swanepoel thanks the National Research Foundation of South Africa for financial support. The authors also acknowledge partial support by the Ministry of the Flemish Community (Project BOF 05B01 International Scientific and Technological Cooperation).

References

Abramson, I.S. (1982). On bandwidth variation in kernel estimates - a square root law. *Ann. Statist.* **10**, 1217-1223.

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

Hall, P. & Marron, J.S. (1988). Variable window width kernel estimates of probability densities. *Prob. Theory Rel. Fields* **80**, 37-50.

Hall, P., Hu, T.C. & Marron, J.S. (1995). Improved variable window kernel estimates of probability densities. *Ann. Statist.* **23**, 1-10.

Jones, M.C., Linton, O. & Nielsen, J.P. (1995). A simple bias reduction method for density estimation. *Biometrika* **82**, 327-338.

Jones, M.C. & Signorini, D.F. (1997). A comparison of higher-order bias kernel density estimators. *J. Amer. Statist. Soc.* **92**, 1063-1073.

Ruppert, D. & Cline, D.B.H. (1994). Bias reduction in kernel density estimation by smoothed empirical transformations. *Ann. Statist.* **22**, 185-210.

Samiuddin, M. & El-Sayyad, G.M. (1990). On nonparametric kernel density estimates. *Biometrika* **77**, 865-874.

Silverman, B.W. (1986). *Density Estimation for Statistics and Data Analysis.* London: Chapman and Hall.

Swanepoel, J.W.H. & Van Graan, F.C. (2005). A new kernel distribution function estimator based on a nonparametric transformation of the data. *Scand. J. Statist.* (under revision).