Computational Statistics (2006) 21:589–601
DOI 10.1007/s00180-006-0016-x

ORIGINAL PAPER

A bias reducing technique in kernel distribution function estimation

Choongrak Kim · Sungsoo Kim · Mira Park · Hakbae Lee

Published online: 4 November 2006

Abstract In this paper we suggest a bias reducing technique in kernel distribution function estimation. In fact, it uses a convex combination of three kernel estimators, and it turned out that the bias has been reduced to the fourth power of the bandwidth, while the bias of the kernel distribution function estimator has the second power of the bandwidth. Also, the variance of the proposed estimator remains at the same order as the kernel distribution function estimator. Numerical results based on simulation studies show this phenomenon, too.

Keywords Bandwidth · Bias · Convex combination · Kernel

1 Introduction

As an estimator of distribution function, empirical distribution function estimator for the complete data or the Kaplan–Meier estimator (Kaplan and Meier 1958) for the censored data is widely used. But, these estimators are step functions, and therefore, they have undesirable properties. To overcome these disadvantages, smoothing versions of them are often used. Among them

C. Kim (✉) · S. Kim
Department of Statistics, Pusan National University, Pusan 609-735, South Korea
e-mail: crkim@pusan.ac.kr

M. Park
School of Medicine, Eulji University, Taejeon 301-832, South Korea

H. Lee
Department of Applied Statistics, Yonsei University, Seoul 120-749, South Korea

kernel smoothing is most widely used because it is easy to derive and has good properties. Kernel smoothing has received a lot of attention in density estimation. Good references in this area are Silverman (1986) and Wand and Jones (1995). However, results in kernel distribution function estimation are relatively few. Theoretical properties of kernel distribution function estimator have been investigated by Nadaraya (1964), Azzalini (1981) and Reiss (1981). Bandwidth selection in kernel distribution function estimator have been suggested by several authors. Among them, Sarda (1993) suggested a 'leave-one-out' method, Altman and Leger (1995) suggested a 'plug-in' method, and Bowman et al. (1998) suggested a 'cross-validation' method.

In this paper we study a bias reducing technique in kernel distribution function estimator by using a convex combination technique of skewed estimators. Convex combination technique to reduce bias in nonparametric regression or density estimation has been used by several authors. Choi and Hall (1998) made bias reduction in local linear regression, Cheng et al. (2000) used skewing method for two parameter locally parametric density estimation. Recently, Kim et al. (2003) showed that the convex combination technique of skewed estimator can be regarded as a fourth order kernel estimator and a new version of the generalized jackknifing approach by Schucany and Sommers (1977). In this paper, we show that the bias of the proposed estimator has $O(h^4)$, while the bias of the kernel distribution function estimator has $O(h^2)$, where h is bandwidth. Also, we show that the variance of the proposed estimator remains at the same order as the kernel distribution function estimator. Numerical results based on simulation studies show this phenomenon, too.

Notations and theoretical properties of kernel distribution function estimator are introduced in Sect. 2. In Sect. 3 the proposed estimator is given and its bias and variance are computed. In Sect. 4, simulation studies are done to see numerical performance of the proposed estimator, and compare it with the kernel distribution function estimator. Finally, concluding remarks are given in Sect. 5.

2 Kernel distribution function estimator

Let $X_1, \ldots, X_n$ be a random sample from a distribution with an unknown density $f(\cdot)$ and distribution function $F(\cdot)$, which we wish to estimate. The kernel estimators of f and F at x are

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

and

$$\hat{F}(x) = \frac{1}{n} \sum_{i=1}^{n} W\left(\frac{x - X_i}{h}\right) \tag{2}$$

respectively, where h is the bandwidth, K is a kernel function, and

$$W(x) = \int_{-\infty}^{x} K(t)\mathrm{d}t$$

is a cumulative kernel function. Assume that K is symmetric, $\int K = 1$, and has a compact support $[-1, 1]$. Let

$$\mu_i = \int_{-1}^{1} t^i K(t)\mathrm{d}t, \quad i = 1, 2, 3, 4. \tag{3}$$

In fact, $\mu_1 = \mu_3 = 0$ since K is symmetric. It is easy to show the following useful relationship;

$$\int_{-1}^{1} W(t)K(t)\mathrm{d}t = \frac{1}{2}, \tag{4}$$

and

$$\int_{-1}^{1} t^i W(t)\mathrm{d}t = \frac{1}{i+1}(1 - \mu_{i+1}), \quad i = 0, 1, 2, 3, 4. \tag{5}$$

Then, by a Taylor expansion and (5) ,

$$\begin{aligned}
E\hat{F}(x) &= \int_{-\infty}^{\infty} W\left(\frac{x-y}{h}\right) f(y)\mathrm{d}y \\
&= \int_{-\infty}^{x-h} 1 \cdot f(y)\mathrm{d}y + \int_{x-h}^{x+h} W\left(\frac{x-y}{h}\right) f(y)\mathrm{d}y + \int_{x+h}^{\infty} 0 \cdot f(y)\mathrm{d}y \\
&= F(x-h) + \int_{-1}^{1} hW(t) f(x-ht)\mathrm{d}t \\
&= F(x) - hf(x) + \frac{1}{2}h^2 f'(x) + o(h^2) \\
&\quad + \int_{-1}^{1} hW(t) \left\{ f(x) - htf'(x) + \frac{1}{2}h^2t^2 f''(x) + o(h^2) \right\} \mathrm{d}t \\
&= F(x) + \frac{1}{2}h^2 f'(x)\mu_2 + o(h^2).
\end{aligned}$$

Therefore, bias of $\hat{F}(x)$ is

$$E\hat{F}(x) - F(x) = \frac{1}{2}h^2 f'(x)\mu_2 + o(h^2), \tag{6}$$

and the variance of $\hat{F}(x)$ can be computed as

$$\mathrm{Var}\left\{\hat{F}(x)\right\} = \frac{1}{n}F(x)(1 - F(x)) + \frac{h}{n}f(x)\left(\int_{-1}^{1} W^2(t)\mathrm{d}t - 1\right) + O\left(\frac{h^2}{n}\right).$$

3 The proposed estimator

It is well known that the kernel density estimator underestimate at peaks and overestimate at troughs. To overcome this problem, Choi and Hall (1998) suggested an estimator based on the convex combination of skewed estimates. This phenomenon also happens in kernel distribution function estimator(see Fig. 1). One method of reducing such biases is using the convex combination, based on the similar idea to Choi and Hall (1998), of three estimates in the following way.

We suggest an estimator of $F(x)$ at x by

$$\tilde{F}(x) = \frac{\lambda_1\hat{F}_1(x) + \hat{F}(x) + \lambda_2\hat{F}_2(x)}{\lambda_1 + 1 + \lambda_2}, \tag{7}$$

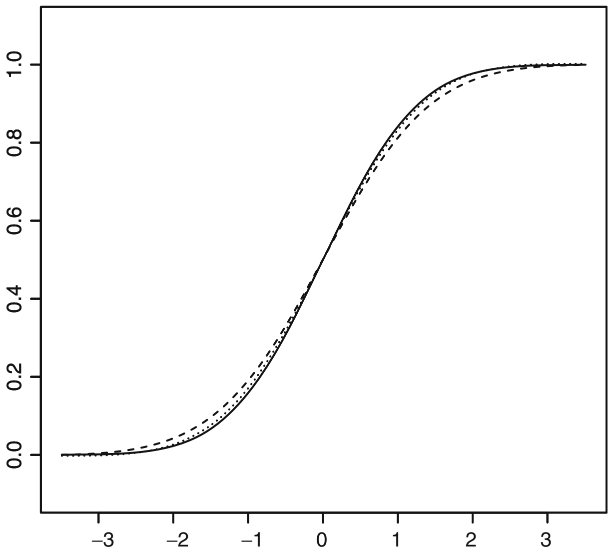

Fig. 1 Kernel distribution function estimator (*dashed line*), the proposed estimator (*dotted line*), and the true standard normal distribution function (*solid line*)

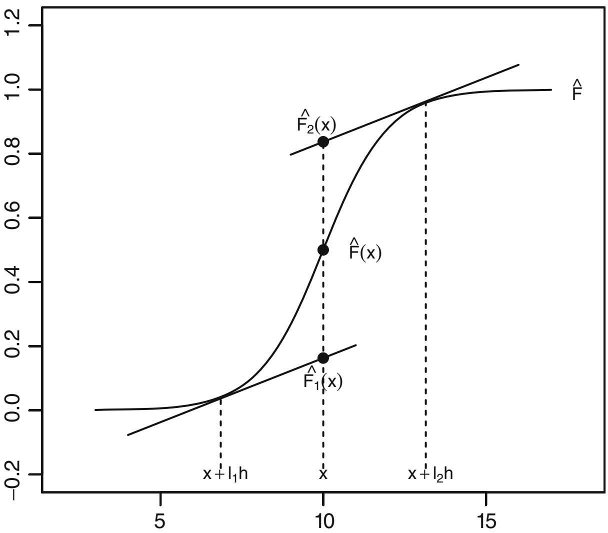

Fig. 2 Convex combination of three kernel estimators $\widehat{F}(x), \widehat{F}_1(x)$ and $\widehat{F}_2(x)$, where $\widehat{F}_j(x) = \widehat{F}(x + l_jh) - l_jh\widehat{f}(x + l_jh), j = 1, 2$

where $\lambda_1,\ \lambda_2 > 0$ are weights,

$$\hat{F}_j(x) = \hat{F}(x + l_jh) - l_jh\hat{f}(x + l_jh), \quad j = 1, 2. \tag{8}$$

$l_1 < 0,\ \ l_2 > 0$ are constants to be determined. The suggested estimator $\tilde{F}(x)$ is a convex combination of $\widehat{F}_1(x),\ \widehat{F}(x)$ and $\widehat{F}_2(x)$. The estimate $\widehat{F}_j(x)$ represents the value, at x, of the tangent line which meets $\widehat{F}(\cdot)$ at $x + l_jh$, as shown in Fig. 2. Choi and Hall (1998), the convex combination based on estimates constructed by three local linear fits in the regression problem. This paper, inspired by Taylor expansions, replaces the local linear fits by linear approximations based on kernel distribution and density estimates in the estimation of distribution function problem. By choosing $\lambda_1 = \lambda_2 = \lambda,\ \ -l_1 = l_2 = l(\lambda)$ with

$$l(\lambda) = \{(1 + 2\lambda)\mu_2/2\lambda\}^{1/2},$$

it can be shown that the bias of $\tilde{F}(x)$ is $O(h^4)$, while that of $\widehat{F}(x)$ is $O(h^2)$. The following theorem shows the bias and the variance of $\tilde{F}(x)$ in detail, and proof of the theorem is given in the Appendix.

Theorem 1 *Assume that F has four bounded, continuous derivatives in a neighborhood of x; that the kernel K is nonnegative, bounded, symmetric, and has compact support* $[-1, 1]$ *; and that $h \to 0$ and $nh \to \infty$. If one takes $\lambda_1 = \lambda_2 = \lambda > 0$ and $-l_1 = l_2 = l(\lambda)$, then*

$$E\{\tilde{F}(x) - F(x)\} = \frac{f'''(x)}{24}\left\{\mu_4 - \frac{3(1 + 6\lambda)}{2\lambda}\mu_2^2\right\}h^4 + o(h^4),$$

$$\mathrm{Var}\{\tilde{F}(x)\} = \frac{1}{n}\frac{2\lambda^2+1}{(2\lambda+1)^2}F(x)(1-F(x)) + \frac{h}{n}f(x)V(\lambda) + O\left(\frac{h^2}{n}\right),$$

where

$$\begin{aligned} V(\lambda) = \frac{1}{(2\lambda+1)^2}\Bigg[(2\lambda^2+1)&\left\{\int_{-1}^{1} W^2(t)\mathrm{d}t + l\int_{-1}^{1} K^2(t)\mathrm{d}t - 1\right\} \\ +2\lambda&\left\{\int_{-1}^{1-l} W(t-l)W(t)\mathrm{d}t + \int_{-1+l}^{1} W(t)W(t+l)\mathrm{d}t \right. \\ &+ \int_{1-l}^{1} (W(t)+\lambda W(t+l))\,\mathrm{d}t - \lambda l \int_{-1}^{-1+2l} K(t)\mathrm{d}t \\ &\left. +\lambda \int_{-1+l}^{1-l} \left(W(t-l)W(t+l) - l^2 K(t-l)K(t+l)\right)\mathrm{d}t\right\}\Bigg]. \end{aligned}$$

Remark 1 Choosing $\lambda = \infty$ in the definition of $\tilde{F}$, we obtain $\tilde{F} = (\widehat{F}_1 + \widehat{F}_2)/2$. We can show that the bias of $\tilde{F}$ with $\lambda = \infty$ is $O(h^3)$. Also, it is obvious that $\tilde{F}$ with $\lambda = 0$ reduces to $\widehat{F}$

Remark 2 For practical use, we need to choose λ and the bandwidth h. Choi and Hall (1998) investigated the effect of λ on the convex combination estimator in local linear regression, and concluded that the estimator is not sensitive to the choice of λ. Based on our limited experience $\tilde{F}(x)$ does not depend seriously on λ. See Sect. 4 for detailed discussions on it. For the choice of h, we may use the cross-validation technique which is widely used in nonparametric density estimation, however, as noted in Sect. 1, other methods are also available.

Remark 3 The order of the leading term, n^{-1}, of the asymptotic variance of $\tilde{F}(x)$ is the same as that of $\hat{F}(x)$. In fact, numerical results from the simulation study in Sect. 4 reveal that variance of each estimator is very close to each other. To see the effect of λ on the variance, we compute the asymptotes at $\lambda = 0$ and $\lambda = \infty$ of the function $V(\lambda)$, and they are

$$V(0) = \int_{-1}^{1} W^2(t)\mathrm{d}t + \sqrt{\mu_2}\int_{-1}^{1} K^2(t)\mathrm{d}t - 1$$

and

$$V(\infty) = \frac{1}{2}\left\{\int_{-1}^{1} W^2(t)\mathrm{d}t + \sqrt{\mu_2}\int_{-1}^{1} K^2(t)\mathrm{d}t - 1 - \sqrt{\mu_2}\int_{-1}^{-1+2\sqrt{\mu_2}} K(t)\mathrm{d}t \right.$$
$$\left. + \int_{-1+\sqrt{\mu_2}}^{1-\sqrt{\mu_2}} \left(W(t-\sqrt{\mu_2})\, W(t+\sqrt{\mu_2}) - \mu_2 K(t-\sqrt{\mu_2})\, K(t+\sqrt{\mu_2})\right) \mathrm{d}t\right\},$$

respectively. $V(\lambda)$ can have a minimum at some point λ depending on the choice of K. The minimizing values of λ are 0.0799 and 0.0769 for the Epanechnikov and uniform kernels, respectively.

Remark 4 The proposed estimator $\tilde{F}(x)$ suffers from a boundary problem at the boundaries of sample. Since the leading term of the bias of $\hat{F}_1(x)$ or $\hat{F}_2(x)$ is $O(h^2)$ (see Appendix), the bias is not worse than $O(h^2)$ even at boundaries. Also, variance remains of order n^{-1} uniformly in the interval of estimation.

4 Simulation studies

To see the numerical performance of the proposed estimator, we made some simulation studies. Based on the standard normal and chi-square distribution with the degree-of-freedom 1 random numbers, we evaluate the mean integrated square error (MISE) for two estimators, the kernel distribution function estimator $\widehat{F}$ and the proposed estimator $\tilde{F}$. In addition to MISE, we like to evaluate the contribution of the bias in MISE since the proposed estimator reduces bias based on the asymptotic results in Theorem 1. To be more specific, let $\hat{\theta}(x)$ be the estimate of $\theta(x)$, where x is a specific point on the support (a, b). Let $\Delta = (b-a)/m$ be the length of m intervals $I_j, j = 1, \ldots, m$, and $\hat{\theta}_i(x)$ be the estimate of $\theta(x)$ based on the ith generated random numbers of size n. Also, let r be the number of replications. Then the Monte Carlo estimator of the MISE is

$$\widehat{\mathrm{MISE}}(\hat{\theta}) = \frac{1}{r}\sum_{i=1}^{r}\sum_{x\in I_1}^{I_m}(\hat{\theta}_i(x) - \theta(x))^2\Delta.$$

The MISE can be written as sum of the IV (integrated variance) and the ISB (integrated squared bias), i.e.,

$$\begin{aligned}\widehat{\mathrm{MISE}}(\hat{\theta}) &= \mathrm{IV} + \mathrm{ISB}\\ &= \frac{1}{r}\sum_{i=1}^{r}\sum_{x\in I_1}^{I_m}(\hat{\theta}_i(x) - \bar{\theta}(x))^2\Delta + \sum_{x\in I_1}^{I_m}(\bar{\theta}(x) - \theta(x))^2\Delta,\end{aligned}$$

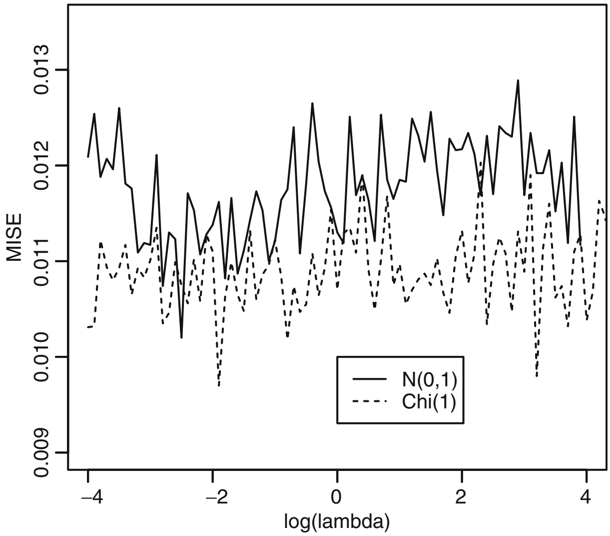

Fig. 3 The MISE in a range of λ for $N(0,1)$ and $\chi^2(1)$

where $\bar{\theta}(x) = \sum_{i=1}^{r} \hat{\theta}_i(x)/r$. In the following, we write MISE$\equiv \widehat{\text{MISE}}(\hat{\theta})$ for notational simplicity.

Thousand replications are done for the sample size $n = 30, 50$ and 100. The Epanechnikov kernel was used in evaluating $\widehat{F}$ and $\tilde{F}$. For $n = 30$, we evaluated the MISE of $\tilde{F}$ for various values of λ, it turned out that the changes in the MISE with respect to λ were negligible for both distributions(see Fig. 3), but the minimum MISE occurred at $\lambda = 0.08$ in $N(0,1)$ case and $\lambda = 0.14$ in $\chi^2(1)$ case. Since the MISE of $\widehat{F}$ and $\tilde{F}$ are function of the bandwidth h, and we choose $h = 1.2$ in $N(0,1)$ and $h = 1.6$ in $\chi^2(1)$ which gives the minimum MISE.

The simulation results are summarized in Table 1. For both distributions, the MISE of two estimators decreases as sample size increases. For each sample size, the MISE of $\tilde{F}$ is smaller than that of $\hat{F}$ in both distributions. Note that the reduction of the MISE of $\tilde{F}$ is mainly due to the bias, and the variance parts for both estimators are very close. These numerical results coincide with the theoretical results in Theorem 1.

5 Concluding remarks

Kernel type smoothing was used in many areas especially in regression and density estimation. However, studies in kernel distribution function estimator are relatively few compared to other areas. In this paper we suggest an estimator for the distribution function using a convex combination of three kernel distribution function estimators. The idea was borrowed from the work by Choi and Hall (1998) in local linear regression. The proposed estimator reduces the bias from $O(h^2)$ to $O(h^4)$, while the variance remains at the same order as

Table 1 The minimum MISE of the kernel distribution function estimator $\widehat{F}$ and the proposed estimator $\tilde{F}$ for $n = 30, 50, 100$

Distribution	n	$\widehat{F}$	$\tilde{F}$
	30	0.0120 (0.0102, 0.0028)	0.0102 (0.0097, 0.0005)
$N(0, 1)$	50	0.0089 (0.0065, 0.0024)	0.0068 (0.0064, 0.0004)
	100	0.0070 (0.0052, 0.0018)	0.0048 (0.0047, 0.0001)
	30	0.0141 (0.0115, 0.0026)	0.0097 (0.0091, 0.0006)
$\chi^2(1)$	50	0.0084 (0.0063, 0.0021)	0.0056 (0.0053, 0.0003)
	100	0.0048 (0.0037, 0.0011)	0.0033 (0.0032, 0.0001)

Values in the parenthesis denote the IV and ISB, respectively

the kernel distribution function estimator. Also, the proposed estimator does not depend seriously on the the weight parameter λ which is used in convex combination. All these theoretical results were validated by the simulation studies. We note that these results can be extended to reducing bias of a smooth version of the Kaplan–Meier estimator when data are censored.

Appendix

Proof of Theorem 1

By a Taylor expansion

$$
\begin{aligned}
& E\widehat{F}(x+lh) \\
&= \int_{-\infty}^{\infty} W\left(\frac{x+lh-y}{h}\right) f(y)\mathrm{d}y \\
&= \int_{-\infty}^{x-(1-l)h} f(y)\mathrm{d}y + \int_{x-(1-l)h}^{x+(1+l)h} W\left(\frac{x+lh-y}{h}\right) f(y)\mathrm{d}y \\
&= F\left(x-(1-l)h\right) + \int_{-1}^{1} W(t) f\left(x-(t-l)h\right) h\mathrm{d}t \\
&= F(x) - (1-l)hf(x) + \frac{1}{2}(1-l)^2h^2f'(x) - \frac{1}{6}(1-l)^3h^3f''(x) \\
&\quad + \frac{1}{24}(1-l)^4h^4f'''(x) + \int_{-1}^{1} hW(t)\left\{f(x) - (t-l)hf'(x)\right. \\
&\quad \left. + \frac{1}{2}(t-l)^2h^2f''(x) - \frac{1}{6}(t-l)^3h^3f'''(x) + o(h^3)\right\} \mathrm{d}t + o(h^4),
\end{aligned}
$$

and by using the following facts;

$$\int (t-l)W(t)\mathrm{d}t = (1-\mu_2)/2 \; - \; l,$$
$$\int (t-l)^2 W(t)\mathrm{d}t = \frac{1}{3} \; - \; (1-\mu_2)l \; + \; l^2,$$
$$\int (t-l)^3 W(t)\mathrm{d}t = (1-\mu_4)/4 \; - \; l \; + \; \frac{3}{2}l^2(1-\mu_2) \; - \; l^3 \, ,$$

we have

$$\begin{aligned} &E\widehat{F}(x+lh) \\ &= F(x) \; + \; lf(x)h \; + \; \frac{1}{2}(l^2 \; + \; \mu_2)f'(x)h^2 \; + \; \frac{1}{6}(l^3 \; + \; 3l\mu_2)f''(x)h^3 \\ &\quad + \frac{1}{24}(l^4 \; + \; 6l^2\mu_2 \; + \; \mu_4)f'''(x)h^4 \; + \; o(h^4). \end{aligned}$$

Also, it is easy to show that

$$\begin{aligned} E\widehat{f}(x+lh) = f(x) + lf'(x)h + \frac{1}{2}(l^2+\mu_2)f''(x)h^2 + \frac{1}{6}(l^3+3l\mu_2)f'''(x)h^3 \\ + \frac{1}{24}(l^4+6l^2\mu_2+\mu_4)f^{(4)}(x)h^4 + o(h^4). \end{aligned}$$

Hence,

$$\begin{aligned} E\left[\widehat{F}(x+lh) - lh\widehat{f}(x+lh)\right] = F(x) + \frac{1}{2}(\mu_2 - l^2)f'(x)h^2 - \frac{1}{3}l^3 f''(x)h^3 \\ + \frac{1}{24}(\mu_4 - 3l^4 - 6l^2\mu_2)f'''(x)h^4 + o(h^4). \end{aligned}$$

Therefore, $E\tilde{F}(x)$ can be computed by letting $l = 0,\ l_1$ and l_2 , and the terms in h^2 and h^3 vanish if and only if

$$\lambda_1 l_1^3 \; + \; \lambda_2 l_2^3 \; = \; 0$$

and

$$\lambda_1(\mu_2 - l_1^2) \; + \; \mu_2 \; + \; \lambda_2(\mu_2 - l_2^2) \; = \; 0.$$

By letting $-l_1 = l_2 = l$, the two equations imply

$$l \equiv l(\lambda) = \sqrt{\frac{2\lambda \; + \; 1}{2\lambda}\mu_2}.$$

Finally, if we substitute $\lambda_1 = \lambda_2 = \lambda$, $-l_1 = l_2 = l = \{\mu_2(2\lambda+1)/2\lambda\}^{1/2}$ for $E\tilde{F}(x)$, then we get the desired result for the bias part. Similar computations, though quite tedious, give the variance part. First, note that

$$\begin{aligned}\mathrm{Var}[\tilde{F}] = \frac{1}{(2\lambda+1)^2}\Big[&\lambda^2\mathrm{Var}(\widehat{F_1}) + \mathrm{Var}(\widehat{F}) + \lambda^2\mathrm{Var}(\widehat{F_2}) \\ &+ 2\lambda\mathrm{Cov}(\widehat{F_1},\widehat{F}) + 2\lambda\mathrm{Cov}(\widehat{F},\widehat{F_2}) + 2\lambda^2\mathrm{Cov}(\widehat{F_1},\widehat{F_2})\Big].\end{aligned}$$

Now, we will compute each term on the right hand side of Var $[\tilde{F}]$ except $\mathrm{Var}(\widehat{F})$ which is given in Sect. 2. Let

$$W_2 = \int_{-1}^{1} W^2(t)\mathrm{d}t, \quad K_2 = \int_{-1}^{1} K^2(t)\mathrm{d}t,$$

then, for the first term, we have

$$\mathrm{Var}(\widehat{F_1}) = \mathrm{Var}[\widehat{F}(x-lh)] + l^2h^2\mathrm{Var}[\widehat{f}(x-lh)] + 2lh\mathrm{Cov}[\widehat{F}(x-lh),\widehat{f}(x-lh)].$$

Now,

$$\begin{aligned}&\mathrm{Var}[\widehat{F}(x-lh)] \\ &\quad= \frac{1}{n}\left[E\left\{W^2\left(\frac{x-lh-y}{h}\right)\right\} - E^2\left\{W\left(\frac{x-lh-y}{h}\right)\right\}\right] \\ &\quad= \frac{1}{n}F(x)\left\{1-F(x)\right\} - \frac{h}{n}f(x)\left\{1+l-W_2-2lF(x)\right\} + O(h^2),\end{aligned}$$

because

$$E\left\{W^2\left(\frac{x-lh-y}{h}\right)\right\} = F(x) - (1+l)hf(x) + hf(x)W_2 + O(h^2)$$

and

$$E^2\left\{W\left(\frac{x-lh-y}{h}\right)\right\} = F^2(x) - 2lhf(x)F(x) + O(h^2).$$

Also, it can be easily shown that

$$\mathrm{Var}[\widehat{f}(x-lh)] = \frac{1}{nh}f(x)K_2 + O(h^2),$$

and

$$\mathrm{Cov}[\widehat{F}(x-lh),\widehat{f}(x-lh)] = \frac{1}{n}\left\{\frac{1}{2} - F(x)\right\} + O(h^2).$$

Therefore,

$$\mathrm{Var}(\widehat{F}_1) = \frac{1}{n}F(x)\{1-F(x)\} + \frac{h}{n}f(x)\left\{W_2 + l^2K_2 - 1\right\} + O(h^2).$$

Similar computations give

$$\mathrm{Var}(\widehat{F}_2) = \frac{1}{n}F(x)\{1-F(x)\} + \frac{h}{n}f(x)\left\{W_2 + l^2K_2 - 1\right\} + O(h^2)$$

since

$$\begin{aligned}&\mathrm{Var}[\widehat{F}(x+lh)]\\ &= \frac{1}{n}\left[E\left\{W^2\left(\frac{x+-y}{h}\right)\right\} - E^2\left\{W\left(\frac{x+lh-y}{h}\right)\right\}\right]\\ &= \frac{1}{n}F(x)\{1-F(x)\} + \frac{h}{n}f(x)\{l-1+W_2-2lF(x)\} + O(h^2).\end{aligned}$$

Now, we have

$$\begin{aligned}&\mathrm{Cov}(\widehat{F}_1,\widehat{F})\\ &= -\frac{1}{n}F^2(x) + \frac{h}{n}f(x)\left\{\int W(t-l)W(t)\mathrm{d}t + \frac{l}{2}\int K(t-l)W(t)\mathrm{d}t\right\} + O(h^2),\end{aligned}$$

$$\begin{aligned}&\mathrm{Cov}(\widehat{F}_2,\widehat{F})\\ &= -\frac{1}{n}F^2(x) + \frac{h}{n}f(x)\left\{\int W(t+l)W(t)\mathrm{d}t - \frac{l}{2}\int K(t+l)W(t)\mathrm{d}t\right\} + O(h^2),\end{aligned}$$

and

$$\begin{aligned}&\mathrm{Cov}(\widehat{F}_1,\widehat{F}_2)\\ &= -\frac{1}{n}F^2(x) + \frac{h}{n}f(x)\left\{\int W(t-l)W(t+l)\mathrm{d}t + l\int K(t-l)W(t+l)dt\right.\\ &\quad \left. -l\int W(t-l)K(t+l)\mathrm{d}t - l^2\int K(t-l)K(t+l)\mathrm{d}t\ \right\} + O(h^2).\end{aligned}$$

By adding up all these terms we have the desired result for the variance.

Acknowledgements This research was supported by Korea Science and Engineering Foundation grant(R14-2003-002-01000-0). The authors thank the editor and referees for their helpful comments which greatly improved the paper.

References

Altman N, Leger C (1995) Bandwidth selection for kernel distribution function estimation. J Stat Plan 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

Bowman A, Hall P, Prvan T (1998) Bandwidth selection for the smoothing of distribution functions. Biometrika 85:799–808

Cheng M-Y, Choi E, Fan J, Hall P (2000) Skewing methods for two-parameter locally parametric density estimation. Bernoulli 6:169–182

Choi E, Hall P (1998) On bias reduction in local linear smoothing. Biometrika 85:333–345

Kaplan EL, Meier P (1958) Nonparametric estimation from incomplete observations. J Am Stat Assoc 53:457–481

Kim C, Kim W, Park B-U (2003) Skewing and generalized jackknifing in kernal density estimation. Commun Stat Theory Methods 32:2153–2162

Nadaraya EA (1964) Some new estimates for distribution functions. Theory Prob Appl 15:497–500

Reiss R-D (1981) Nonparametric estimation of smooth distribution functions. Scand J Stat 8:116–119

Sarda P (1993). Smoothing parameter selection for smooth distribution functions. J Stat Plan Inference 35:65–75

Silverman BW (1986). Density estimation for statistics and data analysis. Chapman and Hall, London

Schucany WR, Sommers JP (1977) Improvement of kernel type estimators. J Am Stat Assoc 72:420–423

Wand MP, Jones MC (1995) Kernel smoothing. Chapman and Hall, London