

ERROR REDUCTION FOR KERNEL DISTRIBUTION FUNCTION ESTIMATORS

Fauzi, Rizky Reza
Graduate School of Mathematics, Kyushu University

Maesono, Yoshihiko
Faculty of Mathematics, Kyushu University

<https://doi.org/10.5109/2232324>

出版情報 : Bulletin of informatics and cybernetics. 49, pp.53-66, 2017-12. Research Association of Statistical Sciences

バージョン :

権利関係 :



**ERROR REDUCTION FOR KERNEL DISTRIBUTION FUNCTION
ESTIMATORS**

by

Rizky Reza FAUZI and Yoshihiko MAESONO

*Reprinted from the Bulletin of Informatics and Cybernetics
Research Association of Statistical Sciences, Vol.49*

FUKUOKA, JAPAN
2017

ERROR REDUCTION FOR KERNEL DISTRIBUTION FUNCTION ESTIMATORS

By

Rizky Reza FAUZI* and Yoshihiko MAESONO†

Abstract

A method to reduce the mean integrated squared error for kernel distribution function estimators is proposed. It can be shown that the asymptotic bias of the proposed method is considerably smaller in the sense of convergence rate than that of the standard kernel distribution function estimator. Even though the rate of convergence of variance does not change, the variance of our proposed method is smaller up to some constants. The idea of this method is using a self-elimination technique between two standard kernel distribution function estimators with different bandwidths, with some helps of exponential and logarithmic expansions. By doing that, vanishing the first term of the asymptotic bias is possible. As a result, mean squared error can be reduced.

Key Words and Phrases: Bias Reduction, Convergence Rate, Distribution Function, Exponential Expansion, Kernel Method, Nonparametric Estimation, Logarithmic Expansion.

1. Introduction

Nonparametric methods are gradually becoming popular in statistical analysis of many fields problems, such as in Economics, Biology, and Actuarial Science. In most cases, this is because of the lack of information about the variable being analysed. Smoothing concerning functions, such as density or cumulative distribution, play a special role in the nonparametric analysis. Knowledge of the density function or distribution function, or their estimates, allows one to characterize the random variable more completely. Especially for the distribution function, we can derive some other characteristics of random variables from that, such as quantiles, survival function, hazard rate, etc.

Let X_1, X_2, \dots, X_n be independently and identically distributed random variables with an absolutely continuous distribution function F_X and a density f_X . The classical nonparametric estimator of F_X has been the empirical distribution function defined by

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

where $I(A)$ denotes the indicator function of a set A . It is obvious that F_n is a step function of height $\frac{1}{n}$ at each observed sample point X_i . When considered as a pointwise

* Graduate School of Mathematics, Kyushu University, Motooka 744, Nishi-ku Fukuoka 819-0395, Japan. E-mail: ma215062@math.kyushu-u.ac.jp

† Faculty of Mathematics, Kyushu University, Motooka 744, Nishi-ku Fukuoka 819-0395, Japan. Email: maesono@math.kyushu-u.ac.jp

estimator of F_X , $F_n(x)$ is an unbiased and strongly consistent estimator of $F_X(x)$. For the global point of view, the Glivenko-Cantelli Theorem implies that

$$\sup \{|F_n(x) - F_X(x)| | x \in \mathbb{R}\} \rightarrow 0 \quad \text{a.s.}$$

as $n \rightarrow \infty$. For details, see section 2.1 of Serfling (1980). However, given the information that F_X is absolutely continuous, it seems to be more appropriate to use a smooth and continuous estimator of F_X rather than the empirical distribution function F_n .

Parzen (1962) and Rosenblatt (1956) introduced kernel density estimator as a smooth and continuous estimator for density function. It is defined as

$$\hat{f}_h(x) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{x - X_i}{h}\right), \quad x \in \mathbb{R}, \quad (2)$$

where K is a function called as kernel and $h > 0$ is called as bandwidth, which is a smoothing parameter that controls the smoothness of \hat{f}_h . It is usually assumed that K is a symmetric (about 0) continuous nonnegative function with $\int_{-\infty}^{\infty} K(v)dv = 1$, as well as $h \rightarrow 0$ and $nh \rightarrow \infty$ when $n \rightarrow \infty$. Since distribution function is actually an integral of density function, this kernel density estimator gave an idea to define a kernel distribution function estimator. Nadaraya (1964) defined it as

$$\hat{F}_h(x) = \frac{1}{n} \sum_{i=1}^n W\left(\frac{x - X_i}{h}\right), \quad x \in \mathbb{R}, \quad (3)$$

where $W(v) = \int_{-\infty}^v K(w)dw$. It is easy to prove that this kernel distribution function estimator is a continuous function, and satisfies all the properties of a distribution function.

Several properties of $\hat{F}_h(x)$ are well known. The almost sure uniform convergence of \hat{F}_h to F_X was proved by Nadaraya (1964), Winter (1973) and Yamato (1973), while Yukich (1989) extended this result to higher dimensions. Watson and Leadbetter (1964) proved the asymptotic normality of $\hat{F}_h(x)$, and Chung-Smirnov property was established by Winter (1979) and Degenhardt (1993), i.e.

$$\limsup_{n \rightarrow \infty} \sqrt{\frac{2n}{\log \log n}} \sup \{|\hat{F}_h(x) - F_X(x)| | t \in \mathbb{R}\} = 1 \quad \text{a.s.}$$

Moreover, it has been shown by several authors that the asymptotic performance of $\hat{F}_h(x)$ is better than that of $F_n(x)$, see Azzalini (1981), Reiss (1981), Falk (1983), Singh *et al.* (1983), Hill (1985), Swanepoel (1988), Shirahata and Chu (1992), and Abdous (1993).

A typical measure of accuracy of $\hat{F}_h(x)$ is the MISE (Mean Integrated Squared Error), defined as

$$MISE(\hat{F}_h) = E \left[\int_{-\infty}^{\infty} \{\hat{F}_h(x) - F_X(x)\}^2 w(x) dx \right], \quad (4)$$

where w is a weight function. Several authors have derived an expression for $MISE(\hat{F}_h)$. In particular, Swanepoel (1988) derived an expression for $w(x) = 1$, Jones (1990) derived

it for $w(x) = \frac{1}{f_X(x)}$, and Altman and Léger (1995) considered a general weight function $w(x)$.

Under the conditions that f_X (the density) has one continuous derivative f'_X , it has been proved by the above-mentioned authors that, as $n \rightarrow \infty$,

$$Bias[\hat{F}_h(x)] = h^2 \frac{f'_X(x)}{2} \int_{-\infty}^{\infty} z^2 K(z) dz + o(h^2), \quad (5)$$

$$Var[\hat{F}_h(x)] = \frac{1}{n} F_X(x)[1 - F_X(x)] - \frac{2h}{n} r_1 f_X(x) + o\left(\frac{h}{n}\right) \quad (6)$$

where $r_1 = \int_{-\infty}^{\infty} yK(y)W(y)dy$. It is easy to show that r_1 is a nonnegative number.

Because we have a relationship

$$MISE(\hat{F}_h) = \int_{-\infty}^{\infty} [Bias^2\{\hat{F}_h(x)\} + Var\{\hat{F}_h(x)\}] dx, \quad (7)$$

if we choose $w(x) = 1$, then it is easy to conclude that

$$\begin{aligned} MISE(\hat{F}_h) &= \frac{h^4}{4} \left[\int_{-\infty}^{\infty} z^2 K(z) dz \right]^2 \int_{-\infty}^{\infty} [f'_X(x)]^2 dx \\ &\quad + \frac{1}{n} \int_{-\infty}^{\infty} F_X(x)[1 - F_X(x)] dx - \frac{2h}{n} r_1 + o\left(h^4 + \frac{h}{n}\right) \end{aligned}$$

provided all the integrals above is finite.

There have been many proposals in the literature for improving the bias property of the standard kernel density estimator. Typically, under sufficient smoothness conditions placed on the underlying density f_X , the bias reduces from $O(h^2)$ to $O(h^4)$, and the variance remains of order $\frac{1}{n}$. Those density estimators that could potentially have greater impact include bias reduction by geometric extrapolation by Terrel and Scott (1980), variable bandwidth kernel estimators by Abramson (1982), variable location estimators by Samiuddin and El-Sayyad (1990), nonparametric transformation estimators by Ruppert and Cline (1994), and multiplicative bias correction estimators by Jones *et al.* (1995). One also could use, of course, a so-called higher order kernel functions, but this has a disadvantage that negative values can be found in the density estimates and distribution function estimates. Hence, using higher order kernel for improving kernel distribution function estimator seems not a very good idea.

Because of the good performances of the method of Terrel and Scott (1980) for density estimator, in section 2 we use a similar idea to improve the standard kernel distribution function estimator. However, instead of using a fixed multiplication factor for the bandwidth, we use a general term for that. It can be shown that the proposed estimator, \tilde{F}_X , has a smaller bias in the sense of convergence rate, that is $O(h^4)$. Furthermore, even though the rate of convergence of variance does not change, the variance of our proposed method is smaller up to some constants. At last, we can conclude that our proposed estimator has improved MISE. Some simulation studies are discussed in the section 3, and detailed proofs can be found in the appendices.

2. MISE Reduction by Geometric Extrapolation

In this section, we shall apply geometric extrapolation method to the kernel distribution function estimator, in order to reduce bias. The idea of reducing bias by geometric extrapolation is doing a self-elimination technique between two standard kernel distribution function estimators with different bandwidths, with some helps of exponential and logarithmic expansions. By doing that, vanishing the h^2 term of the asymptotic bias is possible, and the order of convergence changes to h^4 .

Before we can start our main purpose, we need to impose some assumptions, they are:

- A1.** The kernel K is a nonnegative continuous function, symmetric about 0, and it integrates to 1
- A2.** The integral $\int_{-\infty}^{\infty} w^4 K(x) dw$ is finite
- A3.** The bandwidth $h > 0$ satisfies $h \rightarrow 0$ and $nh \rightarrow \infty$ when $n \rightarrow \infty$
- A4.** The density f_X is three times continuously differentiable, and $f_X^{(4)}$ exists
- A5.** The integrals $\int_{-\infty}^{\infty} \frac{[f_X'(x)]^2}{f_X(x)} dx$ and $\int_{-\infty}^{\infty} f_X'''(x) dx$ are finite.

The first and third assumptions are the usual assumptions for the standard kernel distribution function estimator. Since we shall use exponential and logarithmic expansions, we need A2 and A4 to ensure the validity of our proofs. For the last assumption, it is necessary to make sure we can calculate MISE.

We now ready to begin the explanation about how to modify the standard kernel distribution function estimator and reduce its bias. First, we consider the following theorem.

THEOREM 2.1. *Let $J_h(x) = E[\hat{F}_h(x)]$ and $a(\neq 1)$ be a positive number. Under the assumptions A1 - A4, we have*

$$[J_h(x)]^{t_1} [J_{ah}(x)]^{t_2} = F_X(x) + O(h^4), \quad (8)$$

where $t_1 = \frac{a^2}{a^2-1}$ and $t_2 = -\frac{1}{a^2-1}$.

Remark 1. The function $J_{ah}(x)$ is an expectation of the standard kernel distribution function estimator with ah as the bandwidth, that is, $J_{ah}(x) = E[\hat{F}_{ah}(x)]$, where

$$\hat{F}_{ah}(x) = \frac{1}{n} \sum_{i=1}^n W\left(\frac{x - X_i}{ah}\right), \quad x \in \mathbb{R}.$$

Furthermore, the term after $F_X(x)$ in (8) is in the order h^4 , which is smaller than the order of bias of the standard kernel distribution function estimator. Even though this theorem does not state about a bias of some estimator, it will lead us to the idea to modify the standard kernel distribution function estimator. About the explicit asymptotic formula of $O(h^4)$, we shall present it in the appendices.

The theorem 2.1 gives us an idea to modify kernel distribution function estimator which will have, intuitively, similar property of bias. Hence, we propose a new estimator of distribution function as

$$\tilde{F}_X(x) = [\hat{F}_h(x)]^{\frac{a^2}{a^2-1}} [\hat{F}_{ah}(x)]^{-\frac{1}{a^2-1}}. \quad (9)$$

This idea is actually straightforward, using the fact that the expectation of the operation of two statistics is asymptotically equal (in probability) to the operation of expectation of each statistic. Even though we do not use any concept of converge in probability in our proofs, the idea is still applicable by using Taylor Expansion instead.

Remark 2. As we can see, the number a acts as the second smoothing parameter here, because it controls the smoothness of \hat{F}_{ah} (since it is placed inside the function W), and determines how much the effect of \hat{F}_h and \hat{F}_{ah} as a part of their power. Larger a means the effect of \hat{F}_h is larger for \tilde{F}_X , and vice versa. Furthermore, when $a \rightarrow \infty$, we shall find that $\tilde{F}_X \rightarrow \hat{F}_h$. Oppositely, when a really close to 0, the effect of \hat{F}_h is almost vanished. However, different with bandwidth h , the number a is purely our choice and does not depend on the sample size n . Letting a too close to 0 is not wise, since it acts as a denominator in the argument of function W .

Now, for the bias of our proposed estimator, we have the theorem below. As expected, this leading term in this formulas is actually the same as the explicit form of $O(h^4)$ in theorem 2.1 (see appendices).

THEOREM 2.2. *Under the assumptions A1 - A4, the bias of $\tilde{F}_X(x)$ is given by*

$$Bias[\tilde{F}_X(x)] = h^4 a^2 \frac{b_2^2(x) - 2b_4(x)F_X(x)}{2F_X(x)} + o(h^4) + O\left(\frac{1}{n}\right), \quad (10)$$

where

$$b_2(x) = \frac{f'_X(x)}{2} \int_{-\infty}^{\infty} w^2 K(w) dw \quad \text{and} \quad b_4(x) = \frac{f'''_X(x)}{24} \int_{-\infty}^{\infty} w^4 K(w) dw.$$

Remark 3. The factor $F_X(x)$ gives us some uncertain feelings that this bias may be unbounded in some points of real line. However, even though we did not state it in the theorem, the assumption A5 ensures us that the bias is valid and bounded a.s. on the real line. Furthermore as we expected before, the bias is in the order of h^4 . This order is same as if we use forth order kernel function for the standard kernel distribution function estimator. However, since in some points in real line we may find negative estimates if we use those kind of kernel function, our proposed estimator is more appealing.

Next, we discuss about the property of variance. Interestingly enough, there is no differences between the variance of our proposed estimator and the variance of the standard kernel distribution function estimator, in the sense of convergence order, as stated in the theorem below.

THEOREM 2.3. *Under the assumptions A1 - A4, the variance of $\tilde{F}_X(x)$ is given by*

$$Var[\tilde{F}_X(x)] = \frac{1}{n} F_X(x)[1 - F_X(x)] - \frac{2h(a^4 + 1)r_1 + a^2 r_2}{n(a^2 - 1)^2} f_X(x) + o\left(\frac{h}{n}\right), \quad (11)$$

where

$$r_1 = \int_{-\infty}^{\infty} yK(y)W(y)dy \quad \text{and} \quad r_2 = \int_{-\infty}^{\infty} y \left[K(y)W\left(\frac{y}{a}\right) + \frac{1}{a}W(y)K\left(\frac{y}{a}\right) \right] dy.$$

Remark 4. Actually in many cases, we usually omit the $\frac{h}{n}$ term and just denote it as $O\left(\frac{h}{n}\right)$. However, since the dominant term of the variance of the standard kernel distribution function estimator and our proposed method are same, we need the second order term to compare them. It is easy to show that $\frac{a^2}{(a^2-1)^2}r_2 \geq 0$ and $\frac{a^4+1}{(a^2-1)^2} \geq 1$ when $a < 1$ (which is suggested). Hence, up to some constants, the variance of our proposed estimator is smaller than the kernel distribution function estimator's variance.

As we can see, since both of the bias and the variance of our proposed estimator are smaller, then we can conclude that the MISE of our proposed estimator is smaller than the MISE of the standard kernel distribution function estimator. The following theorem 2.4 below states that clearly.

THEOREM 2.4. *Under the assumptions A1 - A5, the mean integrated square error of \tilde{F}_X is smaller than the MISE of \hat{F}_h . It is given by*

$$\begin{aligned} MISE(\tilde{F}_X) &= h^8 a^4 \int_{-\infty}^{\infty} \left[\frac{b_2^2(x) - 2b_4(x)F_X(x)}{2F_X(x)} \right]^2 dx \\ &\quad + \frac{1}{n} \int_{-\infty}^{\infty} F_X(x)[1 - F_X(x)]dx - \frac{2h(a^4 + 1)r_1 + a^2 r_2}{n(a^2 - 1)^2} + o\left(h^8 + \frac{h}{n}\right). \end{aligned}$$

3. Simulation Study

In this section, we present the results of our simulation study to support the theoretical discussion. In this simulation study, we generated random samples from the standard normal distribution, normal distribution with mean 1 and variance 2, Laplace distribution with mean 0 and scale parameter 1, and Laplace distribution with mean 1 and scale parameter 2. The size of each sample is 50, and we did 1000 repetitions for each case. We ran cross-validation method to choose bandwidth. We calculated AISE (Average Integrated Squared Error) as an estimator for MISE.

Table 1: AISE of Standard Kernel DF and Proposed Method

Estimators	$N(0, 1)$	$N(1, 2)$	$Lap(0, 1)$	$Lap(1, 2)$
standard	0.06523	0.07502	0.09098	0.08511
a=0.01	0.03106	0.034894	0.03043	0.04096
a=0.1	0.03127	0.035002	0.03066	0.04149
a=0.25	0.03199	0.0397	0.04353	0.04488
a=0.5	0.04837	0.0469	0.04499	0.04902
a=0.75	0.04917	0.04761	0.04940	0.04947
a=2	0.05415	0.05017	0.06899	0.06760
a=3	0.05745	0.05032	0.0695	0.06826

In the results, we can see that the proposed method gives us good results, especially if we use smaller a . However, the difference between the result of $a = 0.1$ and $a = 0.01$ is not much.

4. Conclusion

This paper has proposed a modification of the standard kernel distribution function by using geometric extrapolation, which reduce the order of convergence of the bias from $O(h^2)$ to $O(h^4)$. It is demonstrated that the elimination of h^2 can be done by self-elimination technique. The variance is not much influenced, but still we can show that the variance of our proposed estimator is smaller than that of the standard kernel distribution function estimator. As a result, the MISE of the proposed method is reduced. Moreover the result of the simulation study reveal the superior performance of the proposed estimator, especially if we choose a small enough. However, choosing a too small and close to 0 is not wise, since it acts as denominator, and the improvement is not much. Actually, because our estimator has two smoothing parameters, h and a , we cannot use usual methods to choose both of them. Establishing new procedure for choosing h and a simultaneously will be a good study to be considered.

Acknowledgement

The authors gratefully acknowledge *JSPS Grant-in-Aid for Exploratory Research No.15K11995 and Scientific Research(B) No.16H02790*

References

- Abdous, B. (1993). Note on the minimum mean integrated squared error of kernel estimates of a distribution function and its derivatives. *Comm. Statist. Theory Methods* Vol. 22, 603-609.
- Abramson, I. S. (1982). On bandwidth variation in kernel estimates - a square root law. *Ann. Statist.* Vol. 10, 1217-1223.
- Altman, N. and Léger, C. (1995). Bandwidth selection for kernel distribution function estimation. *J. Statist. Plann. Inf.* Vol. 46, 195-214.
- Azzalini, A. (1981). A note on the estimation of a distribution function and quantiles by a kernel method. *Biometrika* Vol. 68, 326-328.
- Degenhardt, H. J. A. (1993). Chung-Smirnov Property for perturbed empirical distribution functions. *Statist. Probab. Lett.* Vol. 16, 97-101.
- Falk, M. (1983). Relative efficiency and deficiency of kernel type estimators of smooth distribution functions. *Statist. Neerl.* Vol. 37, 73-83.
- Hill, P. D. (1985). Kernel estimation of a distribution function. *Comm. Statist. Theory Methods* Vol. 14, 605-620.
- Jones, M. C. (1990). The performance of kernel density functions in kernel distribution function estimation. *Statist. Probab. Lett.* Vol. 9, 129-132.
- Jones, M. C., Linton, O., and Nielsen, J. P. (1995). A simple bias reduction method for density estimation. *Biometrika* Vol. 82, 327-338.
- Nadaraya, E. A. (1964). Some new estimates for distribution functions. *Theory Probability and Applications* Vol. 15, 497-500.
- Parzen, E. (1962). On estimation of a probability density function and mode. *The Annals of Mathematical Statistics* Vol. 32, 1065-1076.

- Reiss, R. D. (1981). Nonparametric estimation of smooth distribution functions. *Scand. J. Statist.* Vol. 8, 116-119.
- Rosenblatt, M. (1956). Remarks on some non-parametric estimates of a density function. *The Annals of Mathematical Statistics* Vol. 27, 832-837.
- Ruppert, D. R. and Cline, D. B. H. (1994). Bias reduction in kernel density estimation by smoothed empirical transformations. *Ann. Statist.* Vol. 22, 185-210.
- Samiuddin, M. and El-Sayyad, G. M. (1990). On nonparametric kernel density estimates. *Biometrika* Vol. 77, 865-874.
- Shirahata, S. and Chu, I. S. (1992). Integrated squared error of kernel type estimator of distribution function. *Ann. Inst. Statist. Math.* Vol. 44, 579-591.
- Singh, R. S., Gasser, T., and Prasad, B. (1983). Nonparametric estimates of distribution functions. *Comm. Statist. Theory Methods* Vol. 12, 2095-2108.
- Sterling, R. J. (1980). *Approximation Theorems of Mathematical Statistics*. Wiley.
- Swanepoel, J. W. H. (1988). Mean integrated squared error properties and optimal kernels when estimating a distribution function. *Comm. Statist. Theory Methods* Vol. 17, 3785-3799.
- Terrel, G. R. and Scott, D. W. (1980). On improving convergence rates for non-negative kernel density estimation. *The Annals of Statistics* Vol. 8, 1160-1163.
- Watson, G. S. and Leadbetter, M. R. (1964). Hazard analysis II. *Sankhya Ser. A* Vol. 26, 101-106.
- Winter, B. B. (1973). Strong uniform consistency of integrals of density estimators. *Can. J. Statist.* Vol. 1, 247-253.
- Winter, B. B. (1979). Convergence rate of perturbed empirical distribution functions. *J. Appl. Probab.* Vol. 16, 163-173.
- Yamato, H. (1973). Uniform convergence of an estimator of a distribution function. *Bulletin of Mathematical Statistics* Vol. 15, 69-78.
- Yukich, J. E. (1989). A note on limit theorems for perturbed empirical processes. *Stochastic Process and Applications* Vol. 33, 163-173.

A Appendices

A1. Proof of Theorem 2.1

Let $J_h(x) = E[\widehat{F}_h(x)]$, and extend the expansion until h^4 term. In detail,

$$\begin{aligned}
J_h(x) &= \int_{-\infty}^{\infty} W\left(\frac{x-v}{h}\right) f_X(v) dv \\
&= W\left(\frac{x-v}{h}\right) F_X(v) \Big|_{-\infty}^{\infty} + \frac{1}{h} \int_{-\infty}^{\infty} F_X(v) H\left(\frac{x-v}{h}\right) dv \\
&= \int_{-\infty}^{\infty} F_X(x-hw) K(w) dw \\
&= \int_{-\infty}^{\infty} K(w) \left[F_X(x) - hw f_X(x) + h^2 f'_X(x) \frac{w^2}{2} \right. \\
&\quad \left. - h^3 f''_X(x) \frac{w^3}{6} + h^4 f'''_X(x) \frac{w^4}{24} - \dots \right] dw \\
&= F_X(x) + h^2 b_2(x) + h^4 b_4(x) + o(h^4) \\
&= F_X(x) \left[1 + h^2 \frac{b_2(x)}{F_X(x)} + h^4 \frac{b_4(x)}{F_X(x)} + o(h^4) \right],
\end{aligned}$$

where $b_2(x) = \frac{f'_X(x)}{2} \int_{-\infty}^{\infty} w^2 K(w) dw$ and $b_4(x) = \frac{f'''_X(x)}{24} \int_{-\infty}^{\infty} w^4 K(w) dw$. By taking a natural logarithm and using its expansion, we have

$$\begin{aligned}
\log J_h(x) &= \log F_X(x) \left[1 + h^2 \frac{b_2(x)}{F_X(x)} + h^4 \frac{b_4(x)}{F_X(x)} + o(h^4) \right] \\
&= \log F_X(x) + \log \left[1 + h^2 \frac{b_2(x)}{F_X(x)} + h^4 \frac{b_4(x)}{F_X(x)} + o(h^4) \right] \\
&= \log F_X(x) + \sum_{k=1}^{\infty} \frac{(-1)^{k-1}}{k} \left[h^2 \frac{b_2(x)}{F_X(x)} + h^4 \frac{b_4(x)}{F_X(x)} + o(h^4) \right]^k \\
&= \log F_X(x) + h^2 \frac{b_2(x)}{F_X(x)} + h^4 \frac{b_4(x)}{F_X(x)} + o(h^4) \\
&\quad - \frac{1}{2F_X^2(x)} [h^4 b_2^2(x) + 2h^6 b_2(x) b_4(x) + h^8 b_4^2(x) + o(h^8)] + \dots \\
&= \log F_X(x) + h^2 \frac{b_2(x)}{F_X(x)} + h^4 \frac{2b_4(x)F_X(x) - b_2^2(x)}{2F_X^2(x)} + o(h^4).
\end{aligned}$$

Next, if we define $J_{ah}(x) = E[\widehat{F}_{ah}(x)]$, i.e.

$$\log J_{ah}(x) = \log F_X(x) + a^2 h^2 \frac{b_2(x)}{F_X(x)} + a^4 h^4 \frac{2b_4(x)F_X(x) - b_2^2(x)}{2F_X^2(x)} + o(h^4),$$

we can set up some conditions to eliminate the term h^2 but keep the term $\log F_X(x)$. Since

$$\begin{aligned} \log[J_h(x)]^{t_1} [J_{ah}(x)]^{t_2} &= t_1 \log J_h(x) + t_2 \log J_{ah}(x) \\ &= (t_1 + t_2) \log F_X(x) + (t_1 + a^2 t_2) h^2 \frac{b_2(x)}{F_X(x)} \\ &\quad + (t_1 + a^4 t_2) h^4 \frac{2b_4(x)F_X(x) - b_2^2(x)}{2F_X^2(x)} + o(h^4), \end{aligned}$$

the conditions we need are $t_1 + t_2 = 1$ and $t_1 + a^2 t_2 = 0$. It is obvious that the solutions are $t_1 = \frac{a^2}{a^2-1}$ and $t_2 = -\frac{1}{a^2-1}$, and we get

$$\log[J_h(x)]^{\frac{a^2}{a^2-1}} [J_{ah}(x)]^{-\frac{1}{a^2-1}} = \log F_X(x) - h^4 a^2 \frac{2b_4(x)F_X(x) - b_2^2(x)}{2F_X^2(x)} + o(h^4).$$

To neutralize the natural logarithmic function, we can use exponential function and its expansion, then we have

$$\begin{aligned} [J_h(x)]^{\frac{a^2}{a^2-1}} [J_{ah}(x)]^{-\frac{1}{a^2-1}} &= \exp[\log\{J_h(x)\}^{\frac{a^2}{a^2-1}} \{J_{ah}(x)\}^{-\frac{1}{a^2-1}}] \\ &= \exp\left[\log F_X(x) - h^4 a^2 \frac{2b_4(x)F_X(x) - b_2^2(x)}{2F_X^2(x)} + o(h^4)\right] \\ &= \exp[\log F_X(x)] \exp\left[-h^4 a^2 \frac{2b_4(x)F_X(x) - b_2^2(x)}{2F_X^2(x)} + o(h^4)\right] \\ &= F_X(x) \sum_{k=0}^{\infty} \frac{(-1)^k}{k!} \left[h^4 a^2 \frac{2b_4(x)F_X(x) - b_2^2(x)}{2F_X^2(x)} + o(h^4)\right]^k \\ &= F_X(x) \left[1 + h^4 a^2 \frac{b_2^2(x) - 2b_4(x)F_X(x)}{2F_X^2(x)} + o(h^4) + \dots\right] \\ &= F_X(x) + h^4 a^2 \frac{b_2^2(x) - 2b_4(x)F_X(x)}{2F_X(x)} + o(h^4) \\ &= F_X(x) + O(h^4). \end{aligned}$$

A2. Proof of Theorem 2.2

In order to investigate the bias of our proposed estimator, we rewrite

$$\widehat{F}_h(x) = J_h(x) + Y$$

and $\widehat{F}_{ah}(x) = J_{ah}(x) + Z$, where Y and Z are random variables with

$$E(Y) = E(Z) = 0,$$

$Var(Y) = Var[\widehat{F}_h(x)]$, and $Var(Z) = Var[\widehat{F}_{ah}(x)]$. These forms are actually reasonable, because of the definition of $J_h(x)$ and $J_{ah}(x)$. Then, by the expansion

$$(1+p)^q = 1 + pq + O(p^2),$$

we have

$$\begin{aligned}
\tilde{F}_X(x) &= [\hat{F}_h(x)]^{\frac{a^2}{a^2-1}} [\hat{F}_{ah}(x)]^{-\frac{1}{a^2-1}} = [J_h(x) + Y]^{\frac{a^2}{a^2-1}} [J_{ah}(x) + Z]^{-\frac{1}{a^2-1}} \\
&= [J_h(x)]^{\frac{a^2}{a^2-1}} [J_{ah}(x)]^{-\frac{1}{a^2-1}} \left[1 + \frac{Y}{J_h(x)} \right]^{\frac{a^2}{a^2-1}} \left[1 + \frac{Z}{J_{ah}(x)} \right]^{-\frac{1}{a^2-1}} \\
&= [J_h(x)]^{\frac{a^2}{a^2-1}} [J_{ah}(x)]^{-\frac{1}{a^2-1}} \left[1 + \frac{a^2}{a^2-1} \frac{Y}{J_h(x)} + O\left\{ \frac{Y^2}{J_h^2(x)} \right\} \right] \\
&\quad \times \left[1 - \frac{1}{a^2-1} \frac{Z}{J_{ah}(x)} + O\left\{ \frac{Z^2}{J_{ah}^2(x)} \right\} \right] \\
&= [J_h(x)]^{\frac{a^2}{a^2-1}} [J_{ah}(x)]^{-\frac{1}{a^2-1}} \\
&\quad \times \left[1 + \frac{a^2}{a^2-1} \frac{Y}{J_h(x)} - \frac{1}{a^2-1} \frac{Z}{J_{ah}(x)} + O\{(Y+Z)^2\} \right] \\
&= [J_h(x)]^{\frac{a^2}{a^2-1}} [J_{ah}(x)]^{-\frac{1}{a^2-1}} + \frac{a^2}{a^2-1} \left[\frac{J_h(x)}{J_{ah}(x)} \right]^{\frac{1}{a^2-1}} Y \\
&\quad - \frac{1}{a^2-1} \left[\frac{J_h(x)}{J_{ah}(x)} \right]^{\frac{a^2}{a^2-1}} Z + O[(Y+Z)^2].
\end{aligned}$$

Hence,

$$\begin{aligned}
E[\tilde{F}_X(x)] &= E \left[\{J_h(x)\}^{\frac{a^2}{a^2-1}} \{J_{ah}(x)\}^{-\frac{1}{a^2-1}} + \frac{a^2}{a^2-1} \left\{ \frac{J_h(x)}{J_{ah}(x)} \right\}^{\frac{1}{a^2-1}} Y \right. \\
&\quad \left. - \frac{1}{a^2-1} \left\{ \frac{J_h(x)}{J_{ah}(x)} \right\}^{\frac{a^2}{a^2-1}} Z + O\{(Y+Z)^2\} \right] \\
&= [J_h(x)]^2 [J_{ah}(x)]^{-1} + \frac{2J_h(x)}{J_{ah}(x)} E(Y) - \left[\frac{J_h(x)}{J_{ah}(x)} \right]^2 E(Z) \\
&\quad + O[E\{(Y+Z)^2\}] \\
&= F_X(x) + h^4 a^2 \frac{b_2^2(x) - 2b_4(x)F_X(x)}{2F_X(x)} + o(h^4) + O\left(\frac{1}{n}\right),
\end{aligned}$$

and the bias is

$$Bias[\tilde{F}_X(x)] = h^4 a^2 \frac{b_2^2(x) - 2b_4(x)F_X(x)}{2F_X(x)} + o(h^4) + O\left(\frac{1}{n}\right).$$

A3. Proof of Theorem 2.3

Before we derive the variance, we need to calculate $\frac{J_h(x)}{J_{ah}(x)}$ first. Once again by using $(1+p)^q = 1 + pq + O(p^2)$, we get

$$\begin{aligned}
 \frac{J_h(x)}{J_{ah}(x)} &= \frac{F_X(x) \left[1 + h^2 \frac{b_2(x)}{F_X(x)} + h^4 \frac{b_4(x)}{F_X(x)} + o(h^4) \right]}{F_X(x) \left[1 + a^2 h^2 \frac{b_2(x)}{F_X(x)} + a^4 h^4 \frac{b_4(x)}{F_X(x)} + o(h^4) \right]} \\
 &= \frac{1 + h^2 \frac{b_2(x)}{F_X(x)} + h^4 \frac{b_4(x)}{F_X(x)} + o(h^4)}{1 + O(h^2)} \\
 &= \left[1 + h^2 \frac{b_2(x)}{F_X(x)} + h^4 \frac{b_4(x)}{F_X(x)} + o(h^4) \right] [1 + O(h^2) + O(h^4)] \\
 &= 1 + h^2 \frac{b_2(x)}{F_X(x)} + h^4 \frac{b_4(x)}{F_X(x)} + o(h^4) + O(h^2) = 1 + O(h^2).
 \end{aligned}$$

The calculation of the variance is

$$\begin{aligned}
 Var[\tilde{F}_X(x)] &= Var \left[\{J_h(x)\}^{\frac{a^2}{a^2-1}} \{J_{ah}(x)\}^{-\frac{1}{a^2-1}} + \frac{a^2}{a^2-1} \left\{ \frac{J_h(x)}{J_{ah}(x)} \right\}^{\frac{1}{a^2-1}} Y \right. \\
 &\quad \left. - \frac{1}{a^2-1} \left\{ \frac{J_h(x)}{J_{ah}(x)} \right\}^{\frac{a^2}{a^2-1}} Z + O\{(Y+Z)^2\} \right] \\
 &= Var \left[\frac{a^2}{a^2-1} \{1 + O(h^2)\}^{\frac{1}{a^2-1}} Y - \frac{1}{a^2-1} \{1 + O(h^2)\}^{\frac{a^2}{a^2-1}} Z \right] \\
 &\quad + O[Var\{(Y+Z)^2\}] \\
 &= Var \left(\frac{a^2}{a^2-1} Y - \frac{1}{a^2-1} Z \right) [1 + O(h^2)] + O \left(\frac{1}{n^2} \right) \\
 &= Var \left(\frac{a^2}{a^2-1} Y - \frac{1}{a^2-1} Z \right) + O \left(\frac{h^2}{n} + \frac{1}{n^2} \right) \\
 &= Var \left[\frac{a^2}{a^2-1} \hat{F}_h(x) - \frac{1}{a^2-1} \hat{F}_{ah}(x) \right] + o \left(\frac{h}{n} \right).
 \end{aligned}$$

Because this is just a variance of linear combination of two standard kernel distribution function estimators, the order of the variance does not change, that is $\frac{1}{n}$. For the explicit formula of the variance, first we calculate

$$\frac{a^4}{(a^2-1)^2} Var[\hat{F}_h(x)] + \frac{1}{(a^2-1)^2} Var[\hat{F}_{ah}(x)] - \frac{2a^2}{(a^2-1)^2} Cov[\hat{F}_h(x), \hat{F}_{ah}(x)].$$

Since we already knew about the formulas of $Var[\hat{F}_h(x)]$ and $Var[\hat{F}_{ah}(x)]$, we only need to take a look to the covariance part, that is

$$\begin{aligned} Cov[\hat{F}_h(x), \hat{F}_{ah}(x)] &= \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n Cov \left[W \left(\frac{x - X_i}{h} \right), W \left(\frac{x - X_j}{ah} \right) \right] \\ &= \frac{1}{n} Cov \left[W \left(\frac{x - X_1}{h} \right), W \left(\frac{x - X_1}{ah} \right) \right] \\ &= \frac{1}{n} \left[E \left\{ W \left(\frac{x - X_1}{h} \right) W \left(\frac{x - X_1}{ah} \right) \right\} \right. \\ &\quad \left. - E \left\{ W \left(\frac{x - X_1}{h} \right) \right\} E \left\{ W \left(\frac{x - X_1}{ah} \right) \right\} \right]. \end{aligned}$$

Because,

$$E \left[W \left(\frac{x - X_1}{h} \right) \right] = E \left[W \left(\frac{x - X_1}{ah} \right) \right] = F_X(x) + O(h^2),$$

we only need to calculate

$$\begin{aligned} E \left[W \left(\frac{x - X_1}{h} \right) W \left(\frac{x - X_1}{ah} \right) \right] &= \int_{-\infty}^{\infty} W \left(\frac{x - v}{h} \right) W \left(\frac{x - v}{ah} \right) f_X(v) dv \\ &= W \left(\frac{x - v}{h} \right) W \left(\frac{x - v}{ah} \right) F_X(v) \Big|_{-\infty}^{\infty} - \int_{-\infty}^{\infty} F_X(x) \left[-\frac{1}{h} K \left(\frac{x - v}{h} \right) \right. \\ &\quad \left. \times W \left(\frac{x - v}{ah} \right) - \frac{1}{ah} W \left(\frac{x - v}{h} \right) K \left(\frac{x - v}{ah} \right) \right] dv \\ &= \frac{1}{h} \int_{-\infty}^{\infty} F_X(x) \left[K \left(\frac{x - v}{h} \right) W \left(\frac{x - v}{ah} \right) + \frac{1}{a} W \left(\frac{x - v}{h} \right) K \left(\frac{x - v}{ah} \right) \right] dv \\ &= \int_{-\infty}^{\infty} F_X(x - yh) \left[K(y) W \left(\frac{y}{a} \right) + \frac{1}{a} W(y) K \left(\frac{y}{a} \right) \right] dy \\ &= \int_{-\infty}^{\infty} [F_X(x) - h y f_X(x) + o(h)] \left[K(y) W \left(\frac{y}{a} \right) + \frac{1}{a} W(y) K \left(\frac{y}{a} \right) \right] dy \\ &= F_X(x) \left[\int_{-\infty}^{\infty} K(y) W \left(\frac{y}{a} \right) dy + \frac{1}{a} \int_{-\infty}^{\infty} W(y) K \left(\frac{y}{a} \right) dy \right] - h f_X(x) r_2 + o(h), \end{aligned}$$

where $r_2 = \int_{-\infty}^{\infty} y \left[K(y) W \left(\frac{y}{a} \right) + \frac{1}{a} W(y) K \left(\frac{y}{a} \right) \right] dy$. For the first term of the right-hand side, we have

$$\begin{aligned} \int_{-\infty}^{\infty} K(y) W \left(\frac{y}{a} \right) dy &= W \left(\frac{y}{a} \right) W(y) \Big|_{-\infty}^{\infty} - \frac{1}{a} \int_{-\infty}^{\infty} W(y) K \left(\frac{y}{a} \right) dy \\ &= 1 - \frac{1}{a} \int_{-\infty}^{\infty} W(y) K \left(\frac{y}{a} \right) dy. \end{aligned}$$

Thus we get

$$E \left[W \left(\frac{x - X_1}{h} \right) W \left(\frac{x - X_1}{ah} \right) \right] = F_X(x) - h f_X(x) r_2 + o(h).$$

As a result, we can show that

$$Cov[\widehat{F}_h(x), \widehat{F}_{ah}(x)] = \frac{1}{n}F_X(x)[1 - F_X(x)] - \frac{h}{n}f_X(x)r_2 + o\left(\frac{h}{n}\right),$$

and then

$$Var[\tilde{F}_X(x)] = \frac{1}{n}F_X(x)[1 - F_X(x)] - \frac{2h}{n} \frac{(a^4 + 1)r_1 + a^2r_2}{(a^2 - 1)^2}f_X(x) + o\left(\frac{h}{n}\right).$$

A4. Proof of Theorem 2.4

Because we have a relationship

$$MISE(\widehat{F}_h) = \int_{-\infty}^{\infty} [Bias^2\{\widehat{F}_h(x)\} + Var\{\widehat{F}_h(x)\}]dx,$$

we obtain

$$\begin{aligned} MISE(\tilde{F}_X) &= h^8 a^4 \int_{-\infty}^{\infty} \left[\frac{b_2^2(x) - 2b_4(x)F_X(x)}{2F_X(x)} \right]^2 dx \\ &\quad + \frac{1}{n} \int_{-\infty}^{\infty} F_X(x)[1 - F_X(x)]dx - \frac{2h}{n} \frac{(a^4 + 1)r_1 + a^2r_2}{(a^2 - 1)^2} + o\left(h^8 + \frac{h}{n}\right), \end{aligned}$$

where

$$b_2(x) = \frac{f'_X(x)}{2} \int_{-\infty}^{\infty} w^2 K(w)dw \quad \text{and} \quad b_4(x) = \frac{f'''_X(x)}{24} \int_{-\infty}^{\infty} w^4 K(w)dw.$$

Received June 2, 2017

Revised August 8, 2017