ON DIRECT KERNEL ESTIMATOR OF DENSITY RATIO

Motoyama, Masanari
National Federation of Agricultural Cooperative Associations

Maesono, Yoshihiko
Faculty of Mathematics, Kyushu University

https://doi.org/10.5109/2233857

出版情報：Bulletin of informatics and cybernetics. 50, pp.27-42, 2018-12. Research Association of Statistical Sciences
バージョン：

ON DIRECT KERNEL ESTIMATOR OF DENSITY RATIO

by

Masanari Motoyama and **Yoshihiko Maesono**

Reprinted from the Bulletin of Informatics and Cybernetics
Research Association of Statistical Sciences, Vol.50

FUKUOKA, JAPAN
2018

Bulletin of Informatics and Cybernetics, Vol. 50, 2018

ON DIRECT KERNEL ESTIMATOR OF DENSITY RATIO

By

Masanari Motoyama* and **Yoshihiko Maesono†**

Abstract

Estimation for a density ratio has an important role in statistical inference. We can use the estimator for testing homogeneity of two samples, detecting change point etc. Let $f(x)$ and $g(x)$ denote probability density functions and $g(x_0) \neq 0$ ($x_0 \in \mathbf{R}$). There are several ways to estimate the density ratio $f(x_0)/g(x_0)$. In this paper we discuss a kernel estimation that is a popular method in nonparametric statistical inference. A naive estimator is constituted from separate estimators of $f(x_0)$ and $g(x_0)$, which we call an indirect estimator. Another estimator is proposed by Ćwik and Mielniczuk (1989), which we call a direct estimator. Extending Ćwik and Mielniczuk (1989)'s method, we propose a new direct estimator, and derive an asymptotic mean squared error. We also prove central limit theorem of the new estimator, and compare mean squared errors of the proposed estimator and the direct estimator by simulation.

1. Introduction

Let $X_1, \cdots, X_m$ and $Y_1, \cdots, Y_n$ denote two independent random samples obtained from two populations with distribution functions $F(x)$ and $G(x)$, and probability density functions $f(x)$ and $g(x)$, respectively. For fixed $x_0 \in \mathbf{R}$, let us consider the estimation of the density ratio $f(x_0)/g(x_0)$ where $g(x_0) \neq 0$. The density ratio is not only related to the likelihood ratio concept, but also has a variety of applications in statistical inferences, like testing equality of two samples, discriminant analysis, estimation of separation measures etc. Let $h = h_n$ be a sequence of positive numbers which converges to 0 and $(m+n)h_n \to \infty$. Let K be a kernel function which satisfies

$$\int_{-\infty}^{\infty} K(u)du = 1.$$

For the sake of simplicity, we assume $m = n$ and use the same kernel $K(t)$ for kernel estimators of $f(x_0)$ and $g(x_0)$, which are given by

$$\widehat{f}(x_0) = \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{x_0 - X_i}{h}\right)$$

and

$$\widehat{g}(x_0) = \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{x_0 - Y_i}{h}\right),$$

* National Federation of Agricultural Cooperative Associations, Otemachi 1-2-1, Chiyoda-ku, Tokyo 100-6832, Japan.
† Faculty of Mathematics, Kyushu University, Motooka 744, Nishi-ku Fukuoka 819-0395, Japan. Email: maesono@math.kyushu-u.ac.jp

respectively. In addition, the kernel estimator of $G(x_0)$ is defined as

$$\widehat{G}(x_0) = \frac{1}{n}\sum_{i=1}^{n} W\left(\frac{x_0 - Y_i}{h}\right),$$

where

$$W(t) = \int_{-\infty}^{t} K(u)du.$$

The kernel estimator was introduced by Fix and Hodges (1951), and Akaike (1954). Rosenblatt (1956) and Parzen (1962) have obtained basic properties of the estimator. Mean integrated squared errors of the kernel density estimators are discussed in many papers. The best order of the bandwidth, which minimizes the mean integrated squared errors of the density estimators, is $O(n^{-1/5})$. There are also many papers which studied bias reduction, bandwidth selection, etc. It is easy to show asymptotic normality of a standardized kernel estimator. First we introduce a indirect and direct estimators of the density ratio $f(x_0)/g(x_0)$. The indirect estimator is defined as $\widehat{f}(x_0)/\widehat{g}(x_0)$. Absava and Nadareishvili (1985) studied asymptotic properties of the indirect estimator and proved strong uniform consistency. Motivated by the fact that $f(x)/g(x)$ is the density function of $G(X_1)$ evaluated at $G(x)$, Ćwik and Mielniczuk (1989) introduced the following direct kernel estimator of $f(x_0)/g(x_0)$

$$\widetilde{\frac{f(x_0)}{g(x_0)}} = \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{G_n(x_0) - G_n(X_i)}{h}\right),$$

where $G_n(x) = \sum_{i=1}^{n} I(Y_i \leq x)/n$ is an empirical distribution function, and $I(\cdot)$ is an indicator function. They discussed its strong uniform consistency, asymptotic normality and applications in discriminant analysis.

These estimators have both advantage and disadvantage. We can see that the indirect estimator has smoothness because of the kernel density estimator is smooth, but the indirect estimator can diverge when the denominator is too small. On the other hand, the direct estimator does not diverge, but it is not smooth because it uses empirical distribution function. Since the empirical distribution is a step function, $\widetilde{\frac{f(x_0)}{g(x_0)}}$ takes discrete values with probability 1. Since $\widehat{G}(\cdot)$ is a smooth function, we propose the following smooth direct estimator

$$\widehat{\frac{f(x_0)}{g(x_0)}} = \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{\widehat{G}(x_0) - \widehat{G}(X_i)}{h}\right).$$

In this paper, we will discuss asymptotic properties of the direct smooth estimator. In Section 2, we derive a mean squared error and the central limit theorem for the smooth direct estimator. In Section 3, we simulate mean squared error of the new estimator $\widehat{\frac{f(x_0)}{g(x_0)}}$ and compare its mean squared errors with the direct estimator $\widetilde{\frac{f(x_0)}{g(x_0)}}$.

2. Mean squared error and central limit theorem

In this section, we will obtain the mean squared errors and show the asymptotic normality of the new estimator. First, we will discuss the mean squared error. For the kernel distribution estimator, we have the following moment evaluations.

THEOREM 2.1. *Assume that* (i)$g'(u)$ *exists and* $|g'(u)| \le M$ *for* $M > 0$ *and* $u \in \mathbf{R}$, (ii) K *is a symmetric around the origin 0 and satisfies* $\int u^2 K(u)du < \infty$. *If* $h = O(n^{-d})(0 < d < 1/2)$, *then for* $p \ge 2$ *we have*

$$E\{|\widehat{G}(x_0) - G(x_0)|^p\} = O(h^{2p}) + O\left(\frac{1}{n^{p/2}}\right).$$

PROOF. Since $\widehat{G}(x_0)$ is a mean of *i.i.d.* random variables, we can apply the moment evaluations of martingales (see Dharmadhikari et al. (1968)). Since $\widehat{G}(x_0)$ has a bias, we consider the following decomposition

$$\begin{aligned} & \widehat{G}(x_0) - G(x_0) \\ = \; & \widehat{G}(x_0) - E\left[W\left(\frac{x_0 - Y_1}{h}\right)\right] + E\left[W\left(\frac{x_0 - Y_1}{h}\right)\right] - G(x_0). \end{aligned}$$

Here we use the symmetric kernel $K(\cdot)$ and then it follows from the discussion of the kernel distribution estimator that

$$E\left[W\left(\frac{x_0 - Y_1}{h}\right)\right] - G(x_0) = -\frac{\mu_2 h^2}{2} g'(x_0) + O(h^3),$$

where $\mu_2 = \int u^2 K(u)du$. Since $W(\cdot)$ is bounded, we get

$$E\left|W\left(\frac{x_0 - Y_1}{h}\right)\right|^p = O(1),$$

and then

$$E\left|W\left(\frac{x_0 - Y_1}{h}\right) - E\left\{W\left(\frac{x_0 - Y_1}{h}\right)\right\}\right|^p = O(1).$$

Therefore, it follows from the moment evaluations of martingales that for $p \ge 2$

$$\begin{aligned} & E\left|\widehat{G}(x_0) - G(x_0)\right|^p \\ \le \; & 2^{p-1} E\left|\widehat{G}(x_0) - E\left[W\left(\frac{x_0 - Y_1}{h}\right)\right]\right|^p + 2^{p-1}\left|E\left[W\left(\frac{x_0 - Y_1}{h}\right)\right] - G(x_0)\right|^p \\ \le \; & cn^{-p/2} E\left|W\left(\frac{x_0 - Y_1}{h}\right) - E(W)\right|^p + 2^{p-1}\left|E\left[W\left(\frac{x_0 - Y_1}{h}\right)\right] - G(x_0)\right|^p \\ = \; & O(n^{-p/2}) + O(h^{2p}). \end{aligned}$$

This completes the proof of the theorem.

Hereafter we assume that the support of the kernel function $K(\cdot)$ is $[-1, 1]$. Let us define

$$r(u) = \frac{f(G^{-1}(u))}{g(G^{-1}(u))}.$$

Then we have the following $AMSE$(asymptotic mean squared error).

THEOREM 2.2. *In addition to the conditions in Theorem 1, we assume that the support of* K *is* $[-1, 1]$, $\int K''(u)du = 0$ *and* $K^{(4)}$ *is bounded. If* $h = O(n^{-d})(1/6 \le d \le 1/4)$ *and* $r^{(4)}(u)$ *is bounded, AMSE of the new estimator is given by*

$$E\left[\left\{\widehat{\frac{f(x_0)}{g(x_0)}} - \frac{f(x_0)}{g(x_0)}\right\}^2\right] = B_1(x_0) + O\left(\frac{1}{n} + h^6\right), \tag{1}$$

where

$$B_1(x_0) = \left\{ \frac{r''(G(x_0))}{2} - \frac{fg''}{2g^2}(x_0) \right\}^2 \mu_2^2 h^4 + \frac{1}{nh} \left\{ \frac{f(x_0)}{g(x_0)} + \frac{f^2(x_0)}{g^3(x_0)} \right\} \int_{-1}^{1} K^2(u)du$$

and

$$r''(G(x_0)) = \frac{1}{g^5(x_0)} \left\{ g(x_0) \left[g(x_0) f''(x_0) - f(x_0) g''(x_0) \right] - 3g'(x_0) \left[g(x_0) f'(x_0) - f(x_0) g'(x_0) \right] \right\}.$$

Proof. See appendix.

Chen et al. (2009) obtained the asymptotic mean squared errors of the indirect and direct estimators as follows:

$$E\left[\left\{ \frac{\widehat{f}(x_0)}{\widehat{g}(x_0)} - \frac{f(x_0)}{g(x_0)} \right\}^2 \right] = B_2(x_0) + O\left(\frac{1}{n} + h^5 \right) \tag{2}$$

and

$$E\left[\left\{ \widetilde{\frac{f(x_0)}{g(x_0)}} - \frac{f(x_0)}{g(x_0)} \right\}^2 \right] = B_3(x_0) + O\left(\sqrt{\frac{\log n}{n^2 h}} + h^2 \sqrt{\frac{\log n}{n}} + \frac{\log n}{n} + h^6 \right), \tag{3}$$

where

$$B_2(x_0) = \left\{ \frac{f''(x_0)}{2g(x_0)} - \frac{fg''}{2g^2}(x_0) \right\}^2 \mu_2^2 h^4 + \frac{1}{nh} \left\{ \frac{f(x_0)}{g^2(x_0)} + \frac{f^2(x_0)}{g^3(x_0)} \right\} \int_{-1}^{1} K^2(u)du$$

and

$$B_3(x_0) = \left\{ \frac{r''(G(x_0))}{2} \right\}^2 \mu_2^2 h^4 + \frac{1}{nh} \left\{ \frac{f(x_0)}{g(x_0)} + \frac{f^2(x_0)}{g^2(x_0)} \right\} \int_{-1}^{1} K^2(u)du.$$

By using asymptotic mean squared errors (1), (2) and (3), we can compare the estimators, theoretically.

Example Let us consider the case that $f(x) = g(x)$ $(x \in \mathbf{R})$, $h = n^{-1/5}$ and $K(u) = 15(1-u^2)^2 I(|u| \leq 1)/16$. It is easy to see that

$$r''(G(x_0)) = 0, \qquad \frac{f''(x_0)}{2g(x_0)} - \frac{fg''}{2g^2}(x_0) = 0, \qquad \mu_2 = \frac{1}{7}, \text{ and } \int_{-1}^{1} K^2(u)du = \frac{5}{7}.$$

If the distribution $F(x)$ is standard normal $N(0,1)$, we get

$$B_1(x_0) = n^{-4/5} \left\{ \frac{(x_0^2 - 1)^2}{196} + \frac{5}{7}\left(1 + \sqrt{2\pi} e^{x_0^2/2} \right) \right\},$$

$$B_2(x_0) = n^{-4/5} \frac{10\sqrt{2\pi}}{7} e^{x_0^2/2},$$

$$B_3(x_0) = n^{-4/5} \frac{10}{7}.$$

Thus we have $B_3(x_0) < B_1(x_0) < B_2(x_0)$ for any $x_0 \in \mathbf{R}$.

Though $\widetilde{\frac{f(x_0)}{g(x_0)}}$ has small asymptotic mean squared error in the above example, it is not smooth estimator. The asymptotic mean squared error of $\widehat{\frac{f(x_0)}{g(x_0)}}$ is smaller than the indirect estimator, and $\widehat{\frac{f(x_0)}{g(x_0)}}$ is smooth. Furthermore, as shown in the next section, when the sample size is small, $\widehat{\frac{f(x_0)}{g(x_0)}}$ has small mean squared errors. As a kernel estimator of the density ratio, we recommend $\widehat{\frac{f(x_0)}{g(x_0)}}$.

For the bandwidth $h = n^{-1/5}$, we have the asymptotic normality of the new estimator as follows.

THEOREM 2.3. *Assume the conditions of Theorem 2. If $h = n^{-1/5}$, we have*

$$\sqrt{nh}\left[\widehat{\frac{f(x_0)}{g(x_0)}} - \frac{f(x_0)}{g(x_0)}\right] \xrightarrow{L} N(\mu_\star, \sigma_\star^2),$$

where

$$\mu_\star = \left\{\frac{r''(G(x_0))}{2} - \frac{fg''}{2g^2}(x_0)\right\}\mu_2,$$

$$\sigma_\star^2 = \left\{\frac{f(x_0)}{g(x_0)} + \frac{f^2(x_0)}{g^3(x_0)}\right\}\int_{-1}^{1} K^2(u)du.$$

PROOF. See appendix.

3. Simulation

We compare the mean squared errors (MSE) of the new estimator $\widehat{\frac{f(x_0)}{g(x_0)}}$ and the direct estimator $\widetilde{\frac{f(x_0)}{g(x_0)}}$ by simulation. Based on 10,000 repetitions, we simulate the mean squared errors when $n = 10$, $h = n^{-1/5}$ and $K(u) = 15(1-u^2)^2 I(|u| \leq 1)/16$ which satisfies the condition $\int K''(u)du = 0$. Since the underlying distribution does not affect the simulation results significantly, we only consider the case that the underlying distribution is a normal $N(\mu, \sigma^2)$. Using the following samples, we simulate MSE of $\widehat{\frac{f(x_0)}{g(x_0)}}$ and the ratio of MSEs of the direct and new estimators.

1. Figure 1 and 2 : $X_i \sim N(0,1)$ and $Y_i \sim N(0,1)$
2. Figure 3 and 4 : $X_i \sim N(0.5,1)$ and $Y_i \sim N(0,1)$
3. Figure 5 and 6 : $X_i \sim N(0,4)$ and $Y_i \sim N(0,1)$
4. Figure 7 and 8 : $X_i \sim N(0,1)$ and $Y_i \sim N(0.5,1)$
5. Figure 9 and 10 : $X_i \sim N(0,1)$ and $Y_i \sim N(0,4)$

In many cases, the simulated MSEs of the new estimator $\widehat{\frac{f(x_0)}{g(x_0)}}$ are smaller than those of the direct estimator. This comes from the fact that the sample size $n = 10$ is small. The new estimator is smooth enough even when the sample size is small, whereas the direct estimator is not smooth. The proposed new estimator is smooth and has small MSE, and then the new estimator is useful.

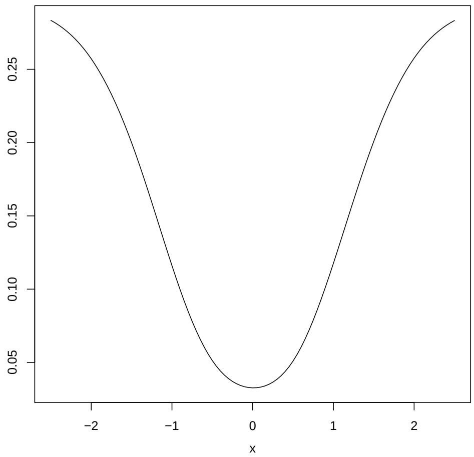

Figure 1: $\text{MSE}\left\{\widehat{\frac{f(x_0)}{g(x_0)}}\right\}$

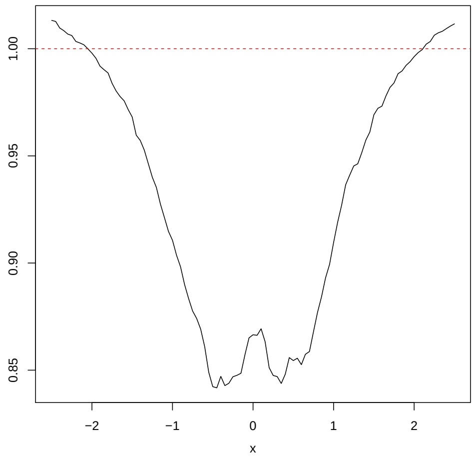

Figure 2: $\text{MSE}\left\{\widehat{\frac{f(x_0)}{g(x_0)}}\right\}/\text{MSE}\left\{\widetilde{\frac{f(x_0)}{g(x_0)}}\right\}$

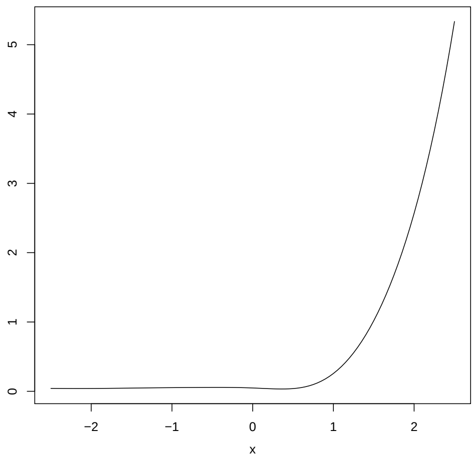

Figure 3: $\text{MSE}\left\{\widehat{\frac{f(x_0)}{g(x_0)}}\right\}$

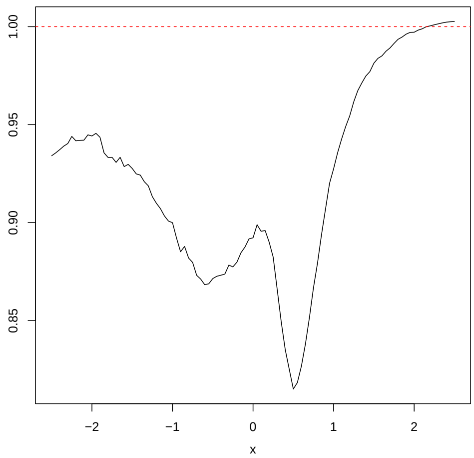

Figure 4: $\text{MSE}\left\{\widehat{\frac{f(x_0)}{g(x_0)}}\right\}/\text{MSE}\left\{\widetilde{\frac{f(x_0)}{g(x_0)}}\right\}$

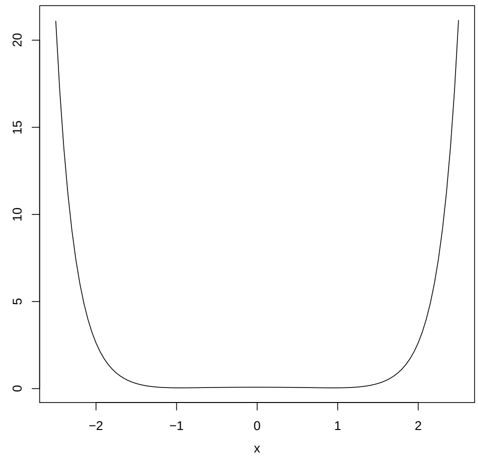

Figure 5: $\mathrm{MSE}\left\{\widehat{\frac{f(x_0)}{g(x_0)}}\right\}$

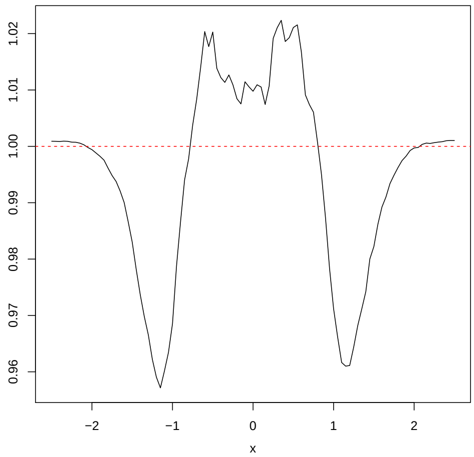

Figure 6: $\mathrm{MSE}\left\{\widehat{\frac{f(x_0)}{g(x_0)}}\right\}/\mathrm{MSE}\left\{\widetilde{\frac{f(x_0)}{g(x_0)}}\right\}$

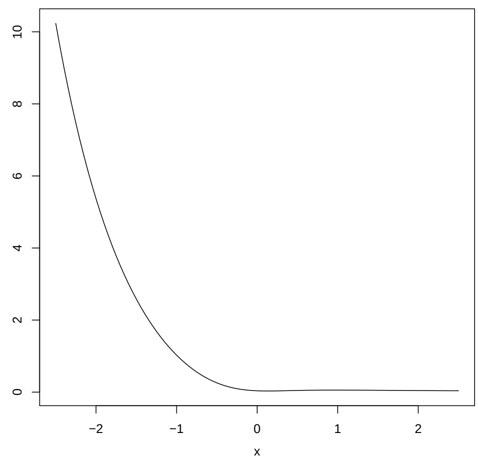

Figure 7: $\mathrm{MSE}\left\{\widehat{\frac{f(x_0)}{g(x_0)}}\right\}$

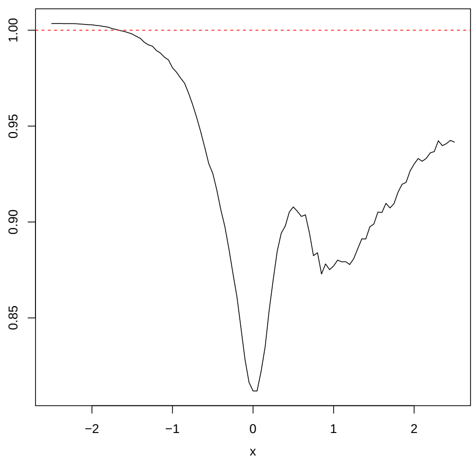

Figure 8: $\mathrm{MSE}\left\{\widehat{\frac{f(x_0)}{g(x_0)}}\right\}/\mathrm{MSE}\left\{\widetilde{\frac{f(x_0)}{g(x_0)}}\right\}$

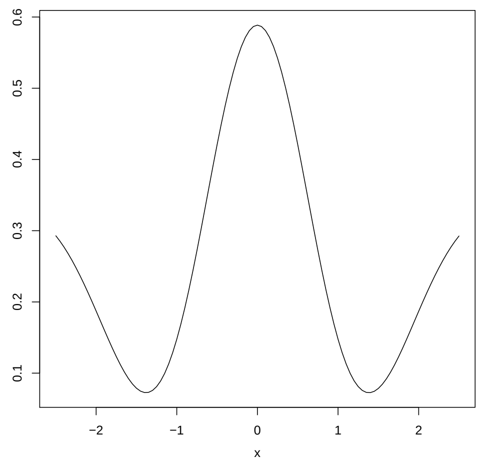

Figure 9: $\text{MSE}\left\{\frac{\widehat{f(x_0)}}{g(x_0)}\right\}$

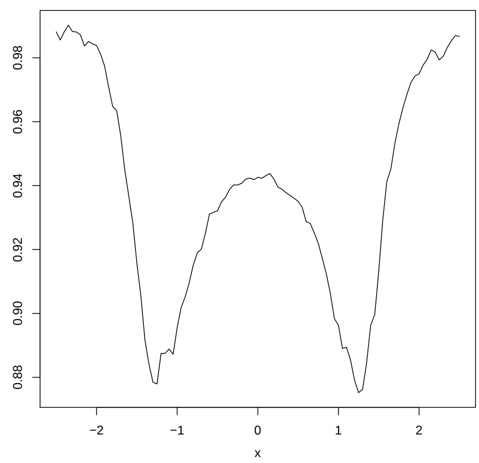

Figure 10: $\text{MSE}\left\{\frac{\widehat{f(x_0)}}{g(x_0)}\right\} / \text{MSE}\left\{\frac{\widetilde{f(x_0)}}{g(x_0)}\right\}$

4. Appendix

Proof of Theorem 2.
By using Taylor expansion we have the following representation

$$
\begin{aligned}
\frac{\widehat{f(x_0)}}{g(x_0)} &= \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{\widehat{G}(x_0)-\widehat{G}(X_i)}{h}\right) \\
&= \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{G(x_0)-G(X_i)}{h}\right) \\
&\quad +\frac{1}{nh}\sum_{i=1}^{n} \frac{1}{h}K'\left(\frac{G(x_0)-G(X_i)}{h}\right)[\{\widehat{G}(x_0)-G(x_0)\}-\{\widehat{G}(X_i)-G(X_i)\}] \\
&\quad +\frac{1}{nh}\sum_{i=1}^{n} \frac{1}{2h^2}K''\left(\frac{G(x_0)-G(X_i)}{h}\right)[\{\widehat{G}(x_0)-G(x_0)\}-\{\widehat{G}(X_i)-G(X_i)\}]^2 \\
&\quad +\frac{1}{nh}\sum_{i=1}^{n} \frac{1}{6h^3}K^{(3)}\left(\frac{G(x_0)-G(X_i)}{h}\right)[\{\widehat{G}(x_0)-G(x_0)\}-\{\widehat{G}(X_i)-G(X_i)\}]^3 \\
&\quad +\frac{1}{nh}\sum_{i=1}^{n} \frac{1}{24h^4}K^{(4)}(\delta)[\{\widehat{G}(x_0)-G(x_0)\}-\{\widehat{G}(X_i)-G(X_i)\}]^4 \\
&= J_1+J_2+J_3+J_4+J_5 \qquad \text{(say)},
\end{aligned}
$$

where δ lies in between $\{\widehat{G}(x_0)-G(x_0)\}/h$ and $\{\widehat{G}(X_i)-G(X_i)\}/h$.

For the first term J_1, by using the transformation $u = [G(x_0) - G(y)]/h$, we get

$$\begin{aligned} E[J_1] &= E\left\{\frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{G(x_0)-G(X_i)}{h}\right)\right\} \\ &= \int_{-\infty}^{\infty} \frac{1}{h} K\left(\frac{G(x_0)-G(y)}{h}\right) f(y)dy \\ &= \int_{[G(x_0)-1]/h}^{G(x_0)/h} K(u)r(G(x_0)-uh)du. \end{aligned}$$

Since the support of $K(\cdot)$ is $[-1,\ 1]$, we get $[G(x_0)-1]/h < -1$ and $G(x_0)/h > 1$ for sufficently small h. Thus we have

$$\int_{[G(x_0)-1]/h}^{G(x_0)/h} K(u)r(G(x_0)-uh)du = \int_{-1}^{1} K(u)r(G(x_0)-uh)du.$$

By using the Taylor expansion, we can show that

$$\begin{aligned} E[J_1] &= r(G(x_0)) + \frac{1}{2}\mu_2 h^2 r''(G(x_0)) + O(h^4) \\ &= \frac{f(x_0)}{g(x_0)} + \frac{1}{2}\mu_2 h^2 r''(G(x_0)) + O(h^4). \end{aligned}$$

Similarly, we can get the following approximation

$$\begin{aligned} Var[J_1] &= \frac{1}{nh^2} Var\left[K\left(\frac{G(x_0)-G(X_1)}{h}\right)\right] \\ &= \frac{1}{nh^2}\left(E\left\{K^2\left(\frac{G(x_0)-G(X_1)}{h}\right)\right\} - \left\{E\left[K\left(\frac{G(x_0)-G(X_1)}{h}\right)\right]\right\}^2\right) \\ &= \frac{1}{nh}\frac{f(x_0)}{g(x_0)}\int_{-1}^{1} K^2(u)du + O\left(\frac{1}{n}\right). \end{aligned}$$

Let us define

$$A_n(x,y) = \{\widehat{G}(x) - G(x)\} - \{\widehat{G}(y) - G(y)\},$$

and then J_2 can be further represented as

$$\begin{aligned}
J_2 &= \frac{1}{nh}\sum_{i=1}^{n}\frac{1}{h}K'\left(\frac{G(x_0)-G(X_i)}{h}\right)A_n(x_0,X_i)\\
&= \frac{1}{h^2}\int K'\left(\frac{G(x_0)-G(y)}{h}\right)A_n(x_0,y)dF_n(y)\\
&= \frac{1}{h^2}\int K'\left(\frac{G(x_0)-G(y)}{h}\right)A_n(x_0,y)d[F_n(y)-F(y)]\\
&\quad +\frac{1}{h^2}\int K'\left(\frac{G(x_0)-G(y)}{h}\right)A_n(x_0,y)dF(y)\\
&= \frac{1}{h^2}\int K'\left(\frac{G(x_0)-G(y)}{h}\right)A_n(x_0,y)d[F_n(y)-F(y)]\\
&\quad +\frac{1}{h^2}\int \frac{f(y)}{g(y)}K'\left(\frac{G(x_0)-G(y)}{h}\right)A_n(x_0,y)dG(y)\\
&= \frac{1}{h^2}\int K'\left(\frac{G(x_0)-G(y)}{h}\right)A_n(x_0,y)d[F_n(y)-F(y)]\\
&\quad +\frac{1}{h^2}\int \left[\frac{f(y)}{g(y)}-\frac{f(x_0)}{g(x_0)}\right]K'\left(\frac{G(x_0)-G(y)}{h}\right)A_n(x_0,y)dG(y)\\
&\quad +\frac{1}{h^2}\int \frac{f(x_0)}{g(x_0)}K'\left(\frac{G(x_0)-G(y)}{h}\right)A_n(x_0,y)dG(y)\\
&= J_{21}+J_{22}+J_{23} \qquad\qquad \text{(say)}.
\end{aligned}$$

We evaluate expectations and variances of the above J_{2k} $(k=1,\cdots,3)$. Since for any function $\alpha(\cdot)$, we have

$$E\left\{\int \alpha(y)d[F_n(y)-F(y)]\right\}=0,$$

where $F_n(y)=\sum_{i=1}^{n} I(X_i\le y)/n$. Then we get

$$\begin{aligned}
E[J_{21}] &= \frac{1}{h^2}E\left\{\int K'\left(\frac{G(x_0)-G(y)}{h}\right)A_n(x_0,y)d[F_n(y)-F(y)]\right\}\\
&= 0
\end{aligned}$$

and if $h=O(n^{-d})$ $(1/6\le d\le 1/4)$, we have

$$\begin{aligned}
Var[J_{21}] &= \frac{1}{h^4}E\biggl\{\int\int K'\left(\frac{G(x_0)-G(y)}{h}\right)K'\left(\frac{G(x_0)-G(z)}{h}\right)\\
&\qquad\qquad \times A_n(x_0,y)A_n(x_0,z)d[F_n(y)-F(y)]d[F_n(z)-F(z)]\biggr\}\\
&= \frac{1}{h^4}E\biggl\{\int\int K'\left(\frac{G(x_0)-G(y)}{h}\right)K'\left(\frac{G(x_0)-G(z)}{h}\right)\\
&\qquad\qquad \times\left\{O(h^4)+O\left(\frac{1}{n}\right)\right\}d[F_n(y)-F(y)]d[F_n(z)-F(z)]\biggr\}\\
&= O\left(\frac{1}{n}\right)+O\left(\frac{1}{n^2h^4}\right)\\
&= O\left(\frac{1}{n}\right).
\end{aligned}$$

Therefore we have

$$E[J_{21}] = 0 \qquad \text{and} \qquad Var[J_{21}] = O\left(\frac{1}{n}\right). \tag{4}$$

Next, we consider J_{22}. By using the Taylor expansion, we get

$$\begin{aligned}
E[J_{22}] &= \frac{1}{h^2}\int\left[\frac{f(y)}{g(y)} - \frac{f(x_0)}{g(x_0)}\right]K'\left(\frac{G(x_0)-G(y)}{h}\right)E\{A_n(x_0,y)\}dG(y) \\
&= \frac{\mu_2}{2}\int\left[\frac{f(y)}{g(y)} - \frac{f(x_0)}{g(x_0)}\right]K'\left(\frac{G(x_0)-G(y)}{h}\right) \\
&\qquad\qquad \times\{g'(x_0) - g'(y) + O(h^2)\}dG(y) \\
&= \frac{\mu_2}{2}h\int_{-1}^{1}\left[\frac{f}{g}(G^{-1}(G(x_0)-uh)) - \frac{f(x_0)}{g(x_0)}\right]K'(u) \\
&\qquad \times\{g'(x_0) - g'(G^{-1}(G(x_0)-uh)) + O(h^2)\}du \\
&= \frac{\mu_2}{2}h\int_{-1}^{1}O(h)uK'(u)\,O(h)udu \\
&= O(h^3).
\end{aligned}$$

Since $|W(y)| \le 1$ and $|G(y)| \le 1$, we get

$$\begin{aligned}
& Var[J_{22}] \\
=\;& E[J_{22}^2] - \{E[J_{22}]\}^2 \\
=\;& E\frac{1}{n^2h^4}\int\int\left[\frac{f(y)}{g(y)} - \frac{f(x_0)}{g(x_0)}\right]\left[\frac{f(z)}{g(z)} - \frac{f(x_0)}{g(x_0)}\right] \\
& \qquad \times K'\left(\frac{G(x_0)-G(y)}{h}\right)K'\left(\frac{G(x_0)-G(z)}{h}\right) \\
& \qquad \times\sum_{i=1}^{n}\left[\left\{W\left(\frac{x_0-Y_i}{h}\right) - G(x_0)\right\} - \left\{W\left(\frac{y-Y_i}{h}\right) - G(y)\right\}\right] \\
& \qquad \times\sum_{j=1}^{n}\left[\left\{W\left(\frac{x_0-Y_j}{h}\right) - G(x_0)\right\} - \left\{W\left(\frac{z-Y_j}{h}\right) - G(z)\right\}\right] \\
& \qquad \times dG(y)dG(z) \\
& - \{E[J_{22}]\}^2 \\
=\;& \frac{1}{nh^4}\int\int\left[\frac{f(y)}{g(y)} - \frac{f(x_0)}{g(x_0)}\right]\left[\frac{f(z)}{g(z)} - \frac{f(x_0)}{g(x_0)}\right] \\
& \qquad \times K'\left(\frac{G(x_0)-G(y)}{h}\right)K'\left(\frac{G(x_0)-G(z)}{h}\right) \\
& \qquad \times E\left[\left[\left\{W\left(\frac{x_0-Y_i}{h}\right) - G(x_0)\right\} - \left\{W\left(\frac{y-Y_i}{h}\right) - G(y)\right\}\right]\right. \\
& \qquad \times\left.\left[\left\{W\left(\frac{x_0-Y_i}{h}\right) - G(x_0)\right\} - \left\{W\left(\frac{z-Y_i}{h}\right) - G(z)\right\}\right]\right] \\
& \qquad \times dG(y)dG(z) \\
& - \frac{1}{n}\{E[J_{22}]\}^2
\end{aligned}$$

$$
\begin{aligned}
&= \frac{1}{nh^4}\left\{h\int_{-1}^{1}\left[\frac{f}{g}(G^{-1}(G(x_0)-uh))-\frac{f(x_0)}{g(x_0)}\right]K'(u)\,O(1)du\right\}^2+O\left(\frac{h^6}{n}\right)\\
&= \frac{1}{nh^2}\left\{\int_{-1}^{1}O(h)uK'(u)\,O(1)du\right\}^2+O\left(\frac{h^6}{n}\right)\\
&= O\left(\frac{1}{n}\right).
\end{aligned}
$$

Therefore, we have

$$
E[J_{22}] = O(h^3) \qquad \text{and} \qquad Var[J_{22}] = O\left(\frac{1}{n}\right). \tag{5}
$$

We proceed to evaluate $E[J_{23}]$. It follows from Lemma 4 of Stute (1984) that

$$
\begin{aligned}
\int K'\left(\frac{G(x_0)-G(y)}{h}\right)dG(y) &= -h\left[K\left(\frac{G(x_0)-G(y)}{h}\right)\right]_{y=-\infty}^{y=\infty}\\
&= -h\left[K\left(\frac{G(x_0)-1}{h}\right)-K\left(\frac{G(x_0)}{h}\right)\right]\\
&= 0,
\end{aligned}
$$

and

$$
\begin{aligned}
&\int\{\widehat{G}(y)-G(y)\}dK\left(\frac{G(x_0)-G(y)}{h}\right)\\
=\ &\left[\{\widehat{G}(y)-G(y)\}K\left(\frac{G(x_0)-G(y)}{h}\right)\right]_{y=-\infty}^{y=\infty}\\
&-\int K\left(\frac{G(x_0)-G(y)}{h}\right)d[\widehat{G}(y)-G(y)]\\
=\ &-\int K\left(\frac{G(x_0)-G(y)}{h}\right)d[\widehat{G}(y)-G(y)]\\
=\ &-\int K\left(\frac{G(x_0)-G(y)}{h}\right)[\widehat{g}(y)-g(y)]dy.
\end{aligned}
$$

Thus we can show that

$$
\begin{aligned}
J_{23} &= \frac{1}{h^2}\int\frac{f(x_0)}{g(x_0)}K'\left(\frac{G(x_0)-G(y)}{h}\right)A_n(x_0,y)dG(y)\\
&= \frac{1}{h^2}\{\widehat{G}(x_0)-G(x_0)\}\frac{f(x_0)}{g(x_0)}\int K'\left(\frac{G(x_0)-G(y)}{h}\right)dG(y)\\
&\quad -\frac{1}{h^2}\frac{f(x_0)}{g(x_0)}\int\{\widehat{G}(y)-G(y)\}K'\left(\frac{G(x_0)-G(y)}{h}\right)dG(y)\\
&= 0-\frac{1}{h^2}\frac{f(x_0)}{g(x_0)}h\int\{\widehat{G}(y)-G(y)\}dK\left(\frac{G(x_0)-G(y)}{h}\right)\\
&= -\frac{1}{h}\frac{f(x_0)}{g(x_0)}\int K\left(\frac{G(x_0)-G(y)}{h}\right)[\widehat{g}(y)-g(y)]dy.
\end{aligned}
$$

By using the Taylor expansion, we have

$$\begin{aligned}
E[J_{23}] &= -\frac{1}{h}\frac{f(x_0)}{g(x_0)}\int K\left(\frac{G(x_0)-G(y)}{h}\right)\left[\frac{1}{2}\mu_2 h^2 g''(y)+O(h^4)\right]dy \\
&= -\frac{1}{2}\mu_2 h^2\frac{f(x_0)}{g(x_0)}\int_{-1}^{1}K(u)\,\frac{g''}{g}(G^{-1}(G(x_0)-uh))du+O(h^4) \\
&= -\frac{1}{2}\mu_2 h^2\frac{fg''}{g^2}(x_0)+O(h^4).
\end{aligned}$$

Similarly, we obtain

$$\begin{aligned}
& Var[J_{23}] \\
=\; & E[J_{23}^2]-\{E[J_{23}]\}^2 \\
=\; & \frac{1}{h^2}\left[\frac{f(x_0)}{g(x_0)}\right]^2\int\int K\left(\frac{G(x_0)-G(y)}{h}\right)K\left(\frac{G(x_0)-G(z)}{h}\right) \\
& \qquad \times E[\widehat{G}(y)-g(y)][\widehat{G}(z)-g(z)]dydz-\{E[J_{23}]\}^2 \\
=\; & \frac{1}{nh^2}\left[\frac{f(x_0)}{g(x_0)}\right]^2\int\int K\left(\frac{G(x_0)-G(y)}{h}\right)K\left(\frac{G(x_0)-G(z)}{h}\right) \\
& \qquad \times E\left[\frac{1}{h}K\left(\frac{y-Y_i}{h}\right)-g(y)\right]\left[\frac{1}{h}K\left(\frac{z-Y_i}{h}\right)-g(z)\right]dydz \\
& -\frac{1}{n}\{E[J_{23}]\}^2 \\
=\; & \frac{1}{nh^2}\left[\frac{f(x_0)}{g(x_0)}\right]^2\int\int K\left(\frac{G(x_0)-G(y)}{h}\right)K\left(\frac{G(x_0)-G(z)}{h}\right) \\
& \qquad \times\left\{\int\frac{1}{h^2}K\left(\frac{y-s}{h}\right)K\left(\frac{z-s}{h}\right)g(s)ds+O(1)+O(h)\right\}dydz \\
& +O\left(\frac{h^4}{n}\right) \\
=\; & \frac{1}{nh^2}\left[\frac{f(x_0)}{g(x_0)}\right]^2\int_{-1}^{1}\int_{-1}^{1}\int K(u)K(v)\frac{1}{g(G^{-1}(G(x_0)-uh))}\frac{1}{g(G^{-1}(G(x_0)-vh))} \\
& \qquad \times K\left(\frac{G^{-1}(G(x_0)-uh)-s}{h}\right)K\left(\frac{G^{-1}(G(x_0)-vh)-s}{h}\right) \\
& \qquad \times g(s)dsdudv+O\left(\frac{1}{n}\right) \\
=\; & \frac{1}{nh^2}\left[\frac{f(x_0)}{g(x_0)}\right]^2\int_{-1}^{1}\int_{-1}^{1}\int K(u)K(v)\left\{\frac{1+uO(h)}{g(x_0)}\right\}\left\{\frac{1+vO(h)}{g(x_0)}\right\} \\
& \qquad \times K\left(\frac{x_0-uO(h)-s}{h}\right)K\left(\frac{x_0-vO(h)-s}{h}\right) \\
& \qquad \times g(s)dsdudv+O\left(\frac{1}{n}\right) \\
=\; & \frac{1}{nh}\left[\frac{f(x_0)}{g(x_0)}\right]^2\int_{-1}^{1}\int_{-1}^{1}\int K(u)K(v)\frac{1}{g^2(x_0)}K^2(t)g(x_0-th)dtdudv+O\left(\frac{1}{n}\right) \\
=\; & \frac{1}{nh}\frac{f^2(x_0)}{g^3(x_0)}\int K^2(t)dt+O\left(\frac{1}{n}\right)
\end{aligned}$$

Therefore, we get

$$E[J_{23}] = -\frac{1}{2}\mu_2 h^2 \frac{fg''}{g^2}(x_0) + O(h^3), \tag{6}$$

$$Var[J_{23}] = \frac{1}{nh}\frac{f^2(x_0)}{g^3(x_0)}\int K^2(t)dt + O\left(\frac{1}{n}\right). \tag{7}$$

Similarly, we can show that

$$\begin{aligned}
& E[J_3] \\
= & \frac{1}{2h^3}E\left\{K''\left(\frac{G(x_0)-G(X_i)}{h}\right)\left\{O(h^4)+O\left(\frac{1}{n}\right)\right\}\right\} \\
= & \frac{1}{2h^3}\left\{O(h^4)+O\left(\frac{1}{n}\right)\right\}\int K''\left(\frac{G(x_0)-G(y)}{h}\right)f(y)dy \\
= & \frac{1}{2h^2}\left\{O(h^4)+O\left(\frac{1}{n}\right)\right\}\int_{-1}^{1} K''(u)r(G(x_0)-uh)du \\
= & \frac{1}{2h^2}\left\{O(h^4)+O\left(\frac{1}{n}\right)\right\}\int_{-1}^{1} K''(u)\left\{r(G(x_0))-uhr'(G(x_0))+u^2O(h^2)\right\}du \\
= & O(h^3).
\end{aligned}$$

Here we use the assumption $\int K''(u)du = 0$. Since $h = O(n^{-d})(1/6 \le d \le 1/4)$, it follows from Theorem 1 that

$$\begin{aligned}
& E[J_3^2] \\
= & \frac{1}{4n^2h^6}E\left\{\sum_{i=1}^{n}K''\left(\frac{G(x_0)-G(X_i)}{h}\right)[\{\widehat{G}(x_0)-G(x_0)\}-\{\widehat{G}(X_i)-G(X_i)\}]^2\right. \\
& \left.\times\sum_{j=1}^{n}K''\left(\frac{G(x_0)-G(X_i)}{h}\right)[\{\widehat{G}(x_0)-G(x_0)\}-\{\widehat{G}(X_j)-G(X_j)\}]^2\right\} \\
= & \frac{1}{4nh^6}E\left\{K''\left(\frac{G(x_0)-G(X_i)}{h}\right)\right\}^2[\{\widehat{G}(x_0)-G(x_0)\}-\{\widehat{G}(X_i)-G(X_i)\}]^4 \\
& +\frac{n-1}{4nh^6}E\left\{K''\left(\frac{G(x_0)-G(X_i)}{h}\right)[\{\widehat{G}(x_0)-G(x_0)\}-\{\widehat{G}(X_i)-G(X_i)\}]^2\right. \\
& \left.\times K''\left(\frac{G(x_0)-G(X_i)}{h}\right)[\{\widehat{G}(x_0)-G(x_0)\}-\{\widehat{G}(X_j)-G(X_j)\}]^2\right\} \\
= & O\left(\frac{1}{nh^6}\right)\left[O(h^8)+O\left(\frac{1}{n^2}\right)\right]+O(1)\left[O(h^8)+O\left(\frac{1}{n^2}\right)\right] \\
= & O\left(\frac{1}{n}\right).
\end{aligned}$$

Thus, we have

$$E[J_3] = O(h^3) \qquad \text{and} \qquad Var[J_3] = O\left(\frac{1}{n}\right). \tag{8}$$

Similarly, by using Theorem 1 we get

$$E[J_4] = O(h^6) + O\left(\frac{1}{n}\right)O(h^2), \qquad Var[J_4] = O\left(\frac{1}{n}\right). \tag{9}$$

and

$$E[J_5] = O(h^3) \qquad \text{and} \qquad Var[J_5] = O\left(\frac{1}{n}\right). \tag{10}$$

Since J_1 is a function of $X_1, \ldots, X_n$ and J_{23} is a function of $Y_1, \ldots, Y_n$, we have $Cov(J_1, J_{23}) = 0$. Using the Cauchy-Schwarz inequality we can show that $Cov(J_1, J_{2k})$ $(k = 1, 2)$ and $Cov(J1, J_\ell)$ $(\ell = 3, 4, 5)$ are all $O(1/n\sqrt{h})$. Combining (4) $\sim$ (10), we can get the proof of Theorem 2.

Proof of Theorem 3

It follows from the definition of each term that

$$\begin{aligned}
\widehat{\frac{f(x_0)}{g(x_0)}} - \frac{f(x_0)}{g(x_0)} &= \frac{1}{h}\int K\left(\frac{G(x_0)-G(y)}{h}\right) d[F_n(y) - F(y)] \\
&\quad + \frac{1}{h}\int K\left(\frac{G(x_0)-G(y)}{h}\right) dF(y) - \frac{f(x_0)}{g(x_0)} \\
&\quad + J_{21} + J_{22} + J_{23} + J_3 + J_4 + J_5.
\end{aligned}$$

From the proof of Theorem 2, we can show that

$$\sqrt{nh}(J_{21} + J_{22} + J_3 + J_4 + J_5) \xrightarrow{P} 0$$

and

$$\sqrt{nh}J_{23} \xrightarrow{P} -\frac{fg''}{2g^2}(x_0).$$

Thus, it is easy to see that

$$\begin{aligned}
&\sqrt{nh}\left[\frac{1}{h}\int K\left(\frac{G(x_0)-G(y)}{h}\right) d[F_n(y) - F(y)] + J_{23}\right] \\
&\qquad\qquad \xrightarrow{L} N\left(-\frac{fg''}{2g^2}(x_0)\mu_2, \sigma_\star^2\right).
\end{aligned}$$

Since

$$\begin{aligned}
&\frac{1}{h}\int K\left(\frac{G(x_0)-G(y)}{h}\right) dF(y) - \frac{f(x_0)}{g(x_0)} \\
= &\frac{1}{h}\int \frac{f(y)}{g(y)} K\left(\frac{G(x_0)-G(y)}{h}\right) dG(y) - \frac{f(x_0)}{g(x_0)} \\
= &\int r(G(x_0) - uh)K(u)du - \frac{f(x_0)}{g(x_0)} \\
= &\ r(G(x_0)) + \frac{r''(G(x_0))}{2}\mu_2 h^2 + O(h^4) - \frac{f(x_0)}{g(x_0)} \\
= &\frac{r''(G(x_0))}{2}\mu_2 h^2 + O(h^4),
\end{aligned}$$

we have

$$\lim_{n\to\infty} \sqrt{nh}\left[\frac{1}{h}\int K\left(\frac{G(x_0)-G(y)}{h}\right) dF(y) - \frac{f(x_0)}{g(x_0)}\right] = \frac{r''(G(x_0))}{2}\mu_2.$$

Finally, by the Slutsky's theorem, we can derive

$$\sqrt{nh}\left[\widehat{\frac{f(x_0)}{g(x_0)}} - \frac{f(x_0)}{g(x_0)}\right] \xrightarrow{L} N(\mu_\star, \sigma_\star^2).$$

Acknowledgement

This work was supported by JSPS KAKENHI Grant number JP16H02790.

References

Absava, R.M. and Nadareishvili, M.M. (1985). Nonparametric estimation of the probability density in the case of the unknown measures in d-dimensional Euclidean space. *Acad. Nauk Gruzin*, Vol. 117, 257-260. In Soobshch.

Akaike, H. (1954). An approximation to the density function*Ann. Inst. Statist. Math.*, Vol. 6, 127-132.

Chen, S.M. and Hsu, Y. S. and Liaw, J. T. (2009). On kernel estimators of density ratio. *Statistics*, Vol. 43, 463-479.

Ćwik, J. and Mielniczuk, J. (1989). Estimating density ratio with application to discriminant analysis. *Comm. Statist. Theory Meth.*, Vol. 18, 3057-3069.

Dharmadhikari, S. W., Fabian, V. and Jogdeo, K. (1968). Bounds on the moments of martingales. *Ann. Math. Statist.*, Vol. 39, 1719-1723.

Fix, E. and Hodges, J. L. (1951). Discriminatory analysis nonparametric discrimination: consistency properties*Report No.4, Project no.21-29-004*.

Parzen, E. (1961). On estimation of a probability density function and mode. *Ann. Math. Statist.*, Vol. 33, 1065-1076.

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

Stute, W. (1984). Asymptotic normality of nearest neighbor regression functions estimates. *Ann. Statist.*, Vol. 12, 917-926.

Received June 8, 2018
Revised August 20, 2018