The Korean Communications
in Statistics Vol. 4, No. 2, 1997
pp. 445 - 452

Improvement of Boundary Bias in Nonparametric Regression via Twicing Technique[1)]

Jaekeun Jo[2)]

Abstract

In this paper, twicing technique for the improvement of asymptotic boundary bias in nonparametric regression is considered. Asymptotic mean squared errors of the nonparametric regression estimators are derived at the boundary region by twicing the Nadaraya-Watson and local linear smoothing. Asymptotic biases of the resulting estimators are of order h^2 and h^4 respectively.

1. Introduction

Let X and Y be random variables which can be modelled by

$$Y = m(X) + \varepsilon, \qquad E\varepsilon = 0 \quad \text{and} \quad \mathrm{Var}(\varepsilon) = v(x),$$

where $m(x)$ and $v(x)$ are smooth functions specifying the conditional mean and variance functions of Y given $X = x$. It is of interest to estimate regression function $m(x) = E(Y|X = x)$ based on a random sample $(X_1, Y_1), \cdots, (X_n, Y_n)$ from (X, Y). Monographs such as Härdle (1990), Wand and Jones (1995) and Fan and Gijbels (1996) provide a good deal of various nonparametric curve fitting procedures. In this paper we assume that $v(x) = \sigma^2$ for simplicity.

Given a curve estimation procedure $\widehat{m}(x) = \widehat{m}(x; x_1, \cdots, x_n; Y_1, \cdots, Y_n)$, twicing estimator $\widetilde{m}(x)$ is defined by the following steps.

Step 1: Compute $\widehat{m}(x)$ and residuals r_i at each x_i, $r_i = Y_i - \widehat{m}(x_i)$, $i = 1, \cdots, n$, from initial smoothing.

Step 2: Apply the smoothing procedure to (x_i, r_i) at x and obtain

---

1) This paper was supported by Kyungsung University Research Fund, 1996.
2) Associate Professor, Department of Computer Science and Statistics, Kyungsung University, Pusan, 608-736, Korea.

$\hat{r}(x) = \hat{m}(x; x_1, \cdots, x_n; r_1, \cdots, r_n)$.

Step 3: Define twicing estimator $\tilde{m}$ at x as $\tilde{m}(x) = \hat{m}(x) + \hat{r}(x)$.

Twicing is a simple modification of original estimator which apply the same smoothing procedure to smooth residuals. For the definition of residual at x_i in Step 1, $\widehat{m_{(-i)}}(x_i)$ (the same version of $\hat{m}$ with all data excluding (x_i, Y_i)) would be more reasonable estimator of m. But since there is no difference in the asymptotic results we will define residual as is given in Step 1 for simplicity. Twicing technique is applied to nonparametric curve estimation by Stuetzle and Mittal (1979) and related research can be found in Abdous (1995) where as a way of constructing classes of higher-order kernel functions, iteration of the twicing procedure is investigated.

In this paper, twicing is applied to the Nadaraya-Watson and local linear smoothing for nonuniform design density case. Mean squared errors of twicing estimators are derived in Theorem 1 and 2 especially when the point where estimation is taken place is at the boundary point. Proofs for the estimation at the interior point is similar and easier than proofs at the boundary and we do not present them here.

2. Asymptotic boundary MSE of twicing estimator

Under the assumptions such as assumptions on page 220 of Kim and Park (1996), we will study asymptotic properties of twicing estimator based on the Nadaraya-Watson and local linear smoothing at the boundary. Unlike estimation at an interior point, estimation near the boundary requires special treatment since the kernel window is devoid of data for smoothing. Here, we consider left boundary case since the treatment of right boundary case is analogous. Suppose that the point at which the estimation is taken place is 0 (we will use notation 0 as a shorthand for 0^+).

Theorem 1. Mean squared error of twicing estimator based on the Nadaraya-Watson smoothing at the left boundary 0 is given by

$$\begin{aligned} \mathrm{MSE}\,(\tilde{m}(0)) &= \left\{\frac{h^2}{2} m''(0)\gamma_2(K^*_{(0)}) + o(h^2)\right\}^2 \\ &\quad + \frac{\sigma^2}{nhf(0)}\gamma_0((K^*_{(0)})^2) + o\left(\frac{1}{nh}\right) \end{aligned}$$

where, for nonnegative integer l, $\gamma_l(K) = \int_0^1 u^l K(u)du$ is the l'th incomplete moment of

K and $K_{(0)}(u)=\frac{1}{\gamma_0(K)}K(u)$ and $K^*(z)=2K(z)-\int K(z-u)K(u)du$ is the 'twicing kernel' of Stuetzle and Mittal (1979).

Proof. Let

$$\begin{aligned} s_l(x) &= n^{-1}\sum_{i=1}^{n}(P_lK)_h(x-x_i) \\ &= (nh)^{-1}\sum_{i=1}^{n}K\left(\frac{x-x_i}{h}\right)\left(\frac{x-x_i}{h}\right)^l \end{aligned} \tag{2.1}$$

and let $(P_lK)_h(x)=\frac{1}{h}K\left(\frac{x}{h}\right)\left(\frac{x}{h}\right)^l$.

The Nadaraya-Watson estimator at a point x is $\widehat{m}(x)=\sum_{i=1}^{n}w_{x,i}Y_i$, where each weight is given by $w_{x,i}=\frac{K_h(x-x_i)}{\sum_{j=1}^{n}K_h(x-x_j)}$. We can see easily that twicing estimator $\widetilde{m}(x)$ is

$$\begin{aligned} \widetilde{m}(x) &= \widehat{m}(x)+\widehat{r}(x) \\ &= 2\sum_{i=1}^{n}w_{x,i}\{Y_i-m(x_i)\}+\sum_{i=1}^{n}w_{x,i}\{m(x_i)-m(x)\} \\ &\quad -\sum_{i=1}^{n}w_{x,i}\{\widehat{m}(x_i)-m(x_i)\}+m(x). \end{aligned} \tag{2.2}$$

If x is left boundary point 0,

$$\begin{aligned} s_l(0) &= n^{-1}\sum_{i=1}^{n}(P_lK)_h(0-x_i) \\ &= f(0)\gamma_l(K)+hf'(0)\gamma_{l+1}(K)+\frac{1}{2}h^2f''(0)\gamma_{l+2}(K) \\ &\quad +O(h^3). \end{aligned} \tag{2.3}$$

From (2.2), we have

$$\begin{aligned} \widetilde{m}(0) &= 2\sum_{i=1}^{n}w_{0,i}\{Y_i-m(x_i)\}+\sum_{i=1}^{n}w_{0,i}\{m(x_i)-m(0)\} \\ &\quad -\sum_{i=1}^{n}w_{0,i}\{\widehat{m}(x_i)-m(x_i)\}+m(0). \end{aligned} \tag{2.4}$$

In this case the second term of (2.4) can be approximated by

$$\sum_{i=1}^{n}w_{0,i}\{m(x_i)-m(0)\} = C_1(m,0)h+C_2(m,f,0)h^2+o(h^2), \tag{2.5}$$

where, for $u=ah$ and $a_l(\alpha)=\int_{-\alpha}^{1}z^lK(z)dz$, (in particular, $a_l(0)=\gamma_l(K)$),

$C_1(m,u)=m'(u)\dfrac{a_1(a)}{a_0(a)}$ and $C_2(m,f,u)$ is a constant factor depending on m,f,u. Since

$$\begin{aligned}&\sum_{i=1}^{n} w_{0,i}\{\widehat{m}(x_i)-m(x_i)\}\\ &=\sum_{i=1}^{n} w_{0,i}\Big\{C_1(m,x_i)h+C_2(m,f,x_i)h^2+\sum_{j=1}^{n} w_{x_i,j}\{Y_j-m(x_j)\}+o(h^2)\Big\},\end{aligned}$$

replacing m by $C_1(m,\cdot)$ and $C_2(m,f,\cdot)$ in (2.5), we get

$$\begin{aligned}\sum_{i=1}^{n} w_{0,i}C_1(m,x_i) &= C_1(m,0)+hC'_1(m,0)\frac{\gamma_1(K)}{\gamma_0(K)}+O(h^2),\\ \sum_{i=1}^{n} w_{0,i}C_2(m,f,x_i) &= C_2(m,f,0)+O(h).\end{aligned}$$

Hence,

$$\begin{aligned}\sum_{i=1}^{n} w_{0,i}\{\widehat{m}(x_i)-m(x_i)\} &= hC_1(m,0)+h^2C'_1(m,0)\frac{\gamma_1(K)}{\gamma_0(K)}\\ &\quad +h^2C_2(m,f,0)+O(h^3).\end{aligned} \tag{2.6}$$

(i) Approximation of bias

From (2.5) and (2.6), the bias of $\widetilde{m}(0)$ can be approximated up to order $O(h^2)$ by

$$-h^2\Big\{C'_1(m,0)\frac{\gamma_0(K)}{\gamma_1(K)}\Big\}=-h^2m''(0)\Big\{\frac{\gamma_1(K)}{\gamma_0(K)}\Big\}^2.$$

For nonnegative integers l, since incomplete moments of $K_{(0)}(u)=\dfrac{1}{\gamma_0(K)}K(u)$ are given by $\gamma_l(K_{(0)})=\dfrac{\gamma_l(K)}{\gamma_0(K)}$, it is easy to show that incomplete moments of twicing kernel $K^*_{(0)}=2K_{(0)}-K_{(0)}*K_{(0)}$ are

$$\begin{aligned}\gamma_1(K^*_{(0)}) &= 0\\ \gamma_2(K^*_{(0)}) &= 2\gamma_2(K_{(0)})-2\gamma_0(K_{(0)})\gamma_2(K_{(0)})+\gamma_1^2(K_{(0)})\\ &= -2\Big\{\frac{\gamma_1(K)}{\gamma_0(K)}\Big\}^2.\end{aligned}$$

Hence, asymptotic bias of $\widetilde{m}(0)$ is given by

$$\frac{h^2}{2}m''(0)\,\gamma_2(K^*_{(0)})+o(h^2).$$

(ii) Approximation of variance

From (2.4), stochastic terms of $\widetilde{m}(0)$ can be written as

$$\sum_{i=1}^{n}\left\{2w_{0,i}-\sum_{j=1}^{n} w_{0,j}\, w_{x_j,i}\right\}(Y_i-m(x_i)), \tag{2.7}$$

where $w_{x_j,i}$ can be approximated by $\frac{K_h(x_j-x_i)}{nf(x_j)a_0(a_j)}+O\left(\frac{h}{n}\right)$. Since $K_{(0)}(u)=\frac{1}{\gamma_0(K)}K(u)$, the sum of products of two weights in (2.7) can be approximated by

$$\begin{aligned}
&\sum_{j=1}^{n} w_{0,j}w_{x_j,i}\\
=\;& \sum_{j=1}^{n}\left\{\frac{K_h(0-x_j)K_h(x_j-x_i)}{n^2f(0)f(x_j)\gamma_0(K)a_0(a_j)}+O\left(\frac{1}{n^2}\right)\right\}\\
=\;& \frac{1}{nf(0)\gamma_0(K)}\left\{n^{-1}\sum_{j=1}^{n}\frac{1}{f(x_j)a_0(a_j)}K_h(x_j-x_i)K_h(0-x_j)\right\}+O\left(\frac{1}{n}\right)\\
=\;& \frac{1}{nhf(0)\gamma_0^2(K)}\left\{\int_0^1 K\left(\frac{0-x_i}{h}-u\right)K(u)du\right\}+O\left(\frac{1}{n^2h^2}+\frac{1}{n}\right)\\
=\;& \frac{1}{nhf(0)\gamma_0^2(K)}(K*K)\left(\frac{0-x_i}{h}\right)+O\left(\frac{1}{n}\right).
\end{aligned}$$

And each weight for Y_i in (2.7) is approximated by

$$\begin{aligned}
2w_{0,i}-\sum_{j=1}^{n} w_{0,j}w_{x_j,i} &= \frac{1}{nhf(0)}(2K_{(0)}-K_{(0)}*K_{(0)})\left(\frac{0-x_i}{h}\right)+O\left(\frac{h}{n}\right)\\
&= \frac{1}{nhf(0)}K^*_{(0)}\left(\frac{0-x_i}{h}\right)+O\left(\frac{h}{n}\right).
\end{aligned}$$

Consequently, variance of $\widetilde{m}(0)$ is approximated by

$$\begin{aligned}
&\sum_{i=1}^{n}\left\{\frac{1}{nhf(0)}K^*_{(0)}(0-x_i\,h)\right\}^2\sigma^2\\
=\;& \frac{1}{nhf^2(0)}\int_0^1 K^*_{(0)}(u)^2 f(0-hu)du\,\sigma^2+O\left(\frac{1}{n^2h^2}\right)\\
=\;& \frac{1}{nhf(0)}\int_0^1 K^*_{(0)}(u)^2du\,\sigma^2+O\left(\frac{1}{n^2h^2}\right)\\
=\;& \frac{1}{nhf(0)}\gamma_0((K^*_{(0)})^2)\sigma^2+o\left(\frac{1}{nh}\right).
\end{aligned}$$

Asymptotic bias and variance of twicing estimator based on the local linear smoothing are given in Theorem 2. The local linear estimator at a point x is $\widehat{m}(x)=\sum_{i=1}^{n} w_{x,i}Y_i$, where for $(P_lK)_h(x)=\frac{1}{h}K\left(\frac{x}{h}\right)\left(\frac{x}{h}\right)^l$ and $s_l(x)=n^{-1}\sum_{i=1}^{n}(P_lK)_h(x-x_i)$,

$$w_{x,i}=\frac{s_2(x)K_h(x-x_i)-s_1(x)(P_1K)_h(x-x_i)}{ns_0(x)s_2(x)-s_1(x)^2} \tag{2.8}$$

is the weight for (x_i, Y_i) at x.

Theorem 2. At the left boundary point 0, mean squared error of twicing estimator based on the local linear smoothing is given by

$$\begin{aligned}\mathrm{MSE}(\tilde{m}(0)) &= \left\{\frac{h^4}{24} m^{(iv)}(0)\gamma_4(K^*_{(1)}) + o(h^4)\right\}^2 \\ &+ \frac{\sigma^2}{nhf(0)}\gamma_0((K^*_{(1)})^2) + o\left(\frac{1}{nh}\right)\end{aligned}$$

where $K_{(1)}(u) = \dfrac{\begin{vmatrix} 1 & \gamma_1(K) \\ u & \gamma_2(K)\end{vmatrix}}{\begin{vmatrix} \gamma_0(K) & \gamma_1(K) \\ \gamma_1(K) & \gamma_2(K)\end{vmatrix}} K(u).$

Proof. Using approximation of $s_i(0)$ in (2.3), we can obtain

$$\begin{aligned}&\sum_{i=1}^{n} w_{0,i}\{m(x_i) - m(0)\} \\ &= \frac{h^2}{2} m''(0)\left(\frac{\gamma_2^2(K) - \gamma_1(K)\gamma_3(K)}{\gamma_0(K)\gamma_2(K) - \gamma_1^2(K)}\right) \\ &+ D_1(m, f, 0)h^3 + D_2(m, f, 0)h^4 + o(h^4)\end{aligned}$$

where $D_1(m, f, u)$ and $D_2(m, f, u)$ are constant factors depending on m, f, u.

(i) Approximation of bias

Using the same procedures in the proofs of Theorem 1, bias of $\tilde{m}(0)$ can be approximated up to $O(h^4)$ by

$$-\frac{h^4}{4} m^{(iv)}(0)\left(\frac{\gamma_2^2(K) - \gamma_1(K)\gamma_3(K)}{\gamma_0(K)\gamma_2(K) - \gamma_1^2(K)}\right)^2.$$

For nonnegative integers l, since incomplete moments of $K_{(1)}(u) = \dfrac{\begin{vmatrix} 1 & \gamma_1(K) \\ u & \gamma_2(K)\end{vmatrix}}{\begin{vmatrix} \gamma_0(K) & \gamma_1(K) \\ \gamma_1(K) & \gamma_2(K)\end{vmatrix}} K(u)$ are given by $\gamma_l(K_{(1)}) = \dfrac{\gamma_2(K)\gamma_l(K) - \gamma_1(K)\gamma_{l+1}(K)}{\gamma_0(K)\gamma_2(K) - \gamma_1^2(K)}$, it is easy to show that the fourth incomplete moment of twicing kernel $K^*_{(1)} = 2K_{(1)} - K_{(1)} * K_{(1)}$ is

$$\begin{aligned}\gamma_4(K^*_{(1)}) &= 2\gamma_4(K_{(1)}) - 2\gamma_0(K_{(1)})\gamma_4(K_{(1)}) - 8\gamma_1(K_{(1)})\gamma_3(K_{(1)}) - 6\gamma_2^2(K_{(1)}) \\ &= -6\gamma_2^2(K_{(1)}) \\ &= -6\left(\frac{\gamma_2^2(K) - \gamma_1(K)\gamma_3(K)}{\gamma_0(K)\gamma_2(K) - \gamma_1^2(K)}\right)^2.\end{aligned}$$

Hence asymptotic bias of $\tilde{m}(0)$ is given by

$$\frac{h^4}{24} m^{(iv)}(0)\gamma_4(K^*_{(1)}) + o(h^4).$$

(ii) Approximation of variance

Stochastic terms of $\tilde{m}(0)$ have been written as

$$\sum_{i=1}^{n}\left\{2w_{0,i} - \sum_{j=1}^{n} w_{0,j}w_{x_j,i}\right\}(Y_i - m(x_i)).$$

Applying (2.3) to the weight $w_{0,i}$ in (2.8) leads to

$$\begin{aligned} w_{0,i} &= \frac{\gamma_2(K)K_h(0-x_i) - \gamma_1(K)(P_1K)_h(0-x_i)}{nf(0)(\gamma_0(K)\gamma_2(K) - \gamma_1^2(K))} + O(n^{-1}) \\ &= \frac{1}{nhf(0)} K_{(1)}\left(\frac{0-x_i}{h}\right) + O(n^{-1}). \end{aligned} \tag{2.9}$$

Using (2.9), the sum of products of two weights can be approximated by

$$\sum_{j=1}^{n} w_{0,j}w_{x_j,i} = \frac{1}{nhf(0)}(K_{(1)} * K_{(1)})\left(\frac{0-x_i}{h}\right) + O(n^{-1}).$$

Thus each weight for Y_i is approximated by

$$\frac{1}{nhf(0)}(2K_{(1)} - K_{(1)} * K_{(1)})\left(\frac{0-x_i}{h}\right).$$

Consequently, variance of $\tilde{m}(0)$ is approximated by

$$\begin{aligned} &\sum_{i=1}^{n}\left\{\frac{1}{nhf(0)} K^*_{(1)}\left(\frac{0-x_i}{h}\right)\right\}^2 \sigma^2 \\ &= \frac{1}{nhf^2(0)} \int_0^1 K^*_{(1)}(u)^2 f(0-hu)du\sigma^2 + O\left(\frac{1}{n^2h^2}\right) \\ &= \frac{1}{nhf(0)} \int_0^1 K^*_{(1)}(u)^2 du\sigma^2 + o\left(\frac{1}{nh}\right) \\ &= \frac{1}{nhf(0)} \gamma_0((K^*_{(1)})^2)\sigma^2 + o\left(\frac{1}{nh}\right). \end{aligned}$$

3. Concluding Remarks

If the point where estimation is taken place is near the boundary such that $x = ah$, we can obtain boundary mean squared error of the twicing estimator based on the Nadaraya-Watson smoothing by

$$\begin{aligned} \text{MSE}(\widetilde{m}(x)) &= \left\{\frac{h^2}{2} m''(x)\gamma_2(K^*_{(0)}) + o(h^2)\right\}^2 \\ &+ \frac{\sigma^2}{nhf(x)} \gamma_0((K^*_{(0)})^2) + o\left(\frac{1}{nh}\right) \end{aligned}$$

where $K_{(0)}(u)$ should be substituted by $\frac{1}{a_0(\alpha)} K(u)$, where $a_l(\alpha) = \int_{-\alpha}^{1} u^l K(u) du$. And for the twicing estimator based on the local linear smoothing,

$$\begin{aligned} \text{MSE}(\widetilde{m}(x)) &= \left\{\frac{h^4}{24} m^{(iv)}(x)\gamma_4(K^*_{(1)}) + o(h^4)\right\}^2 \\ &+ \frac{\sigma^2}{nhf(x)} \gamma_0((K^*_{(1)})^2) + o\left(\frac{1}{nh}\right), \end{aligned}$$

where $K_{(1)}(u)$ should be substituted by $K_{(1)}(u) = \frac{\begin{vmatrix} 1 & a_1(\alpha) \\ u & a_2(\alpha) \end{vmatrix}}{\begin{vmatrix} a_0(\alpha) & a_1(\alpha) \\ a_1(\alpha) & a_2(\alpha) \end{vmatrix}} K(u)$, where $a_l(\alpha) = \int_{-\alpha}^{1} u^l K(u) du$.

At the boundary point, the order of asymptotic bias for the twicing estimator based on the Nadaraya-Watson smoothing is of order h^2 and asymptotic bias for the twicing estimator based on the local linear smoothing is of order h^4. A justification for the good boundary bias of the local linear smoothing is provided by Jones (1993) where generalized jackknifing method is considered to reduce bias order of kernel-type estimators near the boundary.

References

[1] Abdous, B. (1995). Computationally Efficient Classes of Higher-order Kernel Functions. *Canadian Journal of Statistics*. Vol. 23, 21-27.

[2] Fan, J. and Gijbels, I. (1996). *Local Polynomial Modelling and Its Applications*, Chapman & Hall, London.

[3] Härdle, W. (1990). *Applied Nonparametric Regression*, Cambridge University Press, New York.

[4] Jones, M. C. (1993). Simple Boundary Correction for Kernel Density Estimation. *Statistics and Computing*, Vol. 3, 135-146.

[5] Kim, W. C. and Park, B. U. (1996). On a Transformation Technique for Nonparametric Regression. *Journal of the Korean Statistical Society*, Vol. 25, 217-233.

[6] Stuetzle, W. and Mittal, Y. (1979). Some Comments on the Asymptotic Behaviour of Robust Smoothers. In *Smoothing Techniques for Curve Estimation*. Eds. T. Gasser and M. Rosenblatt, 191-195. Splinger Verlag, Berlin.

[7] Wand, M. P. and Jones, M. C. (1995). *Kernel Smoothing*, Chapman & Hall, London.